

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
~~please do not report the images to the~~
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:01 ; Search time 15 seconds
(without alignments)
936.314 million cell updates/sec

Title: US-09-839-136-2
Perfect score: 3096
Sequence: 1 MRPWGSMRWIMILFAMGT.....YKVKREKIEYKPYTPPEADK 575

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PTOUS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3088	99.7	575	3 US-08-913-805A-2	Sequence 2, Appl1
2	3088	99.7	575	4 US-09-442-629-2	Sequence 2, Appl1
3	2949	95.3	575	3 US-08-913-805A-10	Sequence 10, Appl1
4	2949	95.3	575	4 US-09-442-629-10	Sequence 10, Appl1
5	131	4.2	26	3 US-08-913-805A-3	Sequence 3, Appl1
6	131	4.2	26	4 US-09-442-629-3	Sequence 3, Appl1
7	126	4.1	25	3 US-08-913-805A-12	Sequence 12, Appl1
8	126	4.1	25	4 US-09-442-629-12	Sequence 12, Appl1
9	112.5	3.6	1036	4 US-08-968-752B-4	Sequence 4, Appl1
10	112.5	3.6	352	4 US-08-630-915A-12	Sequence 12, Appl1
11	111	3.6	1588	5 PCT-US93-07261-11	Sequence 11, Appl1
12	111	3.6	1663	5 PCT-US93-07261-16	Sequence 16, Appl1
13	110.5	3.6	971	4 US-09-107-149-19	Sequence 19, Appl1
14	110.5	3.6	1074	2 US-08-768-147B-2	Sequence 2, Appl1
15	110.5	3.6	1074	3 US-08-968-752B-2	Sequence 2, Appl1
16	110.5	3.6	1074	4 US-09-107-149-3	Sequence 3, Appl1
17	110	3.6	976	4 US-09-104-324B-4	Sequence 4, Appl1
18	110	3.6	3248	1 US-08-353-700-1	Sequence 1, Appl1
19	110	3.6	3248	5 PCT-US93-16216-1	Sequence 1, Appl1
20	109	3.5	214	5 PCT-US96-08950-2	Sequence 2, Appl1
21	109	3.5	214	5 PCT-US96-09127-2	Sequence 2, Appl1
22	103.5	3.3	677	4 US-08-630-915A-221	Sequence 221, App
23	103.5	3.3	677	4 US-08-836-567-2	Sequence 2, Appl1
24	103.5	3.3	1197	4 US-08-836-567-12	Sequence 12, Appl1
25	103	3.3	1230	2 US-08-968-342C-35	Sequence 35, Appl1
26	102	3.3	251	4 US-08-630-915A-8	Sequence 8, Appl1
27	98	3.2	247	3 US-08-937-271-13	Sequence 13, Appl1

28	98	3.2	343	3 US-08-937-271-17
29	97.5	3.1	676	4 US-09-313-930-2
30	97	3.1	2482	1 US-08-328-254-6
31	97	3.1	2616	6 5206163-3
32	96.5	3.1	546	4 US-08-630-915A-14
33	96.5	3.1	1052	3 US-08-863-118-2
34	96	3.1	1566	2 US-08-687-956A-23
35	95.5	3.1	546	2 US-08-942-423-4
36	95.5	3.1	1052	3 US-08-863-118-1
37	95.5	3.1	1052	4 US-09-377-310-2
38	95	3.1	922	4 US-09-141-206-6
39	94.5	3.1	1053	3 US-08-863-118-3
40	94.5	3.1	1375	4 US-09-722-139-2
41	94	3.0	933	4 US-09-141-206-2
42	94	3.0	933	4 US-09-107-149-17
43	94	3.0	2544	2 US-08-576-626A-32
44	93	3.0	933	4 US-09-107-149-2
45	92	3.0	344	4 US-09-298-886-11

ALIGNMENTS

RESULT 1
US-08-913-805A-2
Sequence 2, Application US/08913805A
Patent No. 6054304
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YAMAGIOANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 35,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear

Sequence 17, Appl
Sequence 2, Appl1
Sequence 6, Appl1
Patent No. 5206163
Sequence 14, Appl
Sequence 2, Appl1
Sequence 23, Appl
Sequence 4, Appl1
Sequence 1, Appl1
Sequence 2, Appl1
Sequence 6, Appl1
Sequence 3, Appl1
Sequence 2, Appl1
Sequence 2, Appl1
Sequence 17, Appl
Sequence 32, Appl
Sequence 2, Appl1
Sequence 11, Appl

MOLECULE TYPE: protein
US-08-913-805A-2

Query Match 99.7%; Score 3088; DB 3; Length 575;
Best Local Similarity 99.8%; Pred. No. 4, 1e-290;
Matches 574; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRPWGSMRWIMILIFAMGTLIFYIGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60
DB 1 MRPWGSMRWIMILIFAMGTLIFYIGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60
QY 61 RRMAGSLRIPEGPIIDOGPASGRVRALEEQPMKAKEOIEYKKOTKNGPKDHEILRRRIE 120
DB 61 RRMAGSLRIPEGPIIDOGPASGRVRALEEQPMKAKEOIEYKKOTKNGPKDHEILRRRIE 120
QY 121 NGAKELMFFLOSLELKKLKNLEGNELORHADEFSLDGHHERSITMDLYLSOTDGAQDMR 180
DB 121 NGAKELMFFLOSLELKKLKNLEGNELORHADEFSLDGHHERSITMDLYLSOTDGAQDMR 180
QY 181 EKEAKDLTELVRRTIYLQNPKDCSKAKKLVCNINKGCGYCOLHHVYCFMAYGTQRT 240
DB 181 EKEAKDLTELVRRTIYLQNPKDCSKAKKLVCNINKGCGYCOLHHVYCFMAYGTQRT 240
QY 241 LALESHNMRYATGMEVFRPVSETCTDRSGSGTHMSGEVKDKNQVVELPIVDSVHPR 300
DB 241 LALESHNMRYATGMEVFRPVSETCTDRSGSGTHMSGEVKDKNQVVELPIVDSVHPR 300
QY 301 PPLPLAVPEDLADRLVRVHGDPVAVMWVSQFYKYLIRPOPMLEKEIEATKKLGFEHPRVI 360
DB 301 PPLPLAVPEDLADRLVRVHGDPVAVMWVSQFYKYLIRPOPMLEKEIEATKKLGFEHPRVI 360
QY 361 GVHVRRTDKVGAAAFHPRIEYTVHVEEDPOLLARRMQYDKKRVLYATDDPALLKEAKTK 420
DB 361 GVHVRRTDKVGAAAFHPRIEYTVHVEEDPOLLARRMQYDKKRVLYATDDPALLKEAKTK 420
QY 421 YPSYEFISDMSISMSAGLNRRYTENSLRGYIIDIHFLSQADFLVCTFSSQVCAVAEIMQ 480
DB 421 YPSYEFISDMSISMSAGLNRRYTENSLRGYIIDIHFLSQADFLVCTFSSQVCAVAEIMQ 480
QY 481 ALHPDASANFRSLDDIYFEGGPAHNAQIAIYHPQPTGEIIPMEPDIIIGVAGNHWDGYP 540
DB 481 ALHPDASANFRSLDDIYFEGGPAHNAQIAIYHPQPTGEIIPMEPDIIIGVAGNHWDGYP 540
QY 541 KGVNRRKLGRTGLPSYKVRKEIETVYKPYTPPEADK 575
DB 541 KGVNRRKLGRTGLPSYKVRKEIETVYKPYTPPEADK 575

RESULT 2
US-09-442-629-2
Sequence 2, Application US/09442629
Patent No. 6291219
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
UOZUMI, Naotumi
SHIBA, Tetsuo
YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC, compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Nordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,629

FILING DATE: 18-NOV-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-0796
TELEFAX: 202-429-1776
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-442-629-2

Query Match 99.7%; Score 3088; DB 4; Length 575;
Best Local Similarity 99.8%; Pred. No. 4, 1e-290;
Matches 574; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRPWGSMRWIMILIFAMGTLIFYIGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60
DB 1 MRPWGSMRWIMILIFAMGTLIFYIGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60
QY 61 RRMAGSLRIPEGPIIDOGPASGRVRALEEQPMKAKEOIEYKKOTKNGPKDHEILRRRIE 120
DB 61 RRMAGSLRIPEGPIIDOGPASGRVRALEEQPMKAKEOIEYKKOTKNGPKDHEILRRRIE 120
QY 121 NGAKELMFFLOSLELKKLKNLEGNELORHADEFSLDGHHERSITMDLYLSOTDGAQDMR 180
DB 121 NGAKELMFFLOSLELKKLKNLEGNELORHADEFSLDGHHERSITMDLYLSOTDGAQDMR 180
QY 181 EKEAKDLTELVRRTIYLQNPKDCSKAKKLVCNINKGCGYCOLHHVYCFMAYGTQRT 240
DB 181 EKEAKDLTELVRRTIYLQNPKDCSKAKKLVCNINKGCGYCOLHHVYCFMAYGTQRT 240
QY 241 LALESHNMRYATGMEVFRPVSETCTDRSGSGTHMSGEVKDKNQVVELPIVDSVHPR 300
DB 241 LALESHNMRYATGMEVFRPVSETCTDRSGSGTHMSGEVKDKNQVVELPIVDSVHPR 300
QY 301 PPLPLAVPEDLADRLVRVHGDPVAVMWVSQFYKYLIRPOPMLEKEIEATKKLGFEHPRVI 360
DB 301 PPLPLAVPEDLADRLVRVHGDPVAVMWVSQFYKYLIRPOPMLEKEIEATKKLGFEHPRVI 360
QY 361 GVHVRRTDKVGAAAFHPRIEYTVHVEEDPOLLARRMQYDKKRVLYATDDPALLKEAKTK 420
DB 361 GVHVRRTDKVGAAAFHPRIEYTVHVEEDPOLLARRMQYDKKRVLYATDDPALLKEAKTK 420
QY 421 YPSYEFISDMSISMSAGLNRRYTENSLRGYIIDIHFLSQADFLVCTFSSQVCAVAEIMQ 480
DB 421 YPSYEFISDMSISMSAGLNRRYTENSLRGYIIDIHFLSQADFLVCTFSSQVCAVAEIMQ 480
QY 481 ALHPDASANFRSLDDIYFEGGPAHNAQIAIYHPQPTGEIIPMEPDIIIGVAGNHWDGYP 540
DB 481 ALHPDASANFRSLDDIYFEGGPAHNAQIAIYHPQPTGEIIPMEPDIIIGVAGNHWDGYP 540
QY 541 KGVNRRKLGRTGLPSYKVRKEIETVYKPYTPPEADK 575
DB 541 KGVNRRKLGRTGLPSYKVRKEIETVYKPYTPPEADK 575

RESULT 3
US-08-913-805A-10
Sequence 10, Application US/08913805A
Patent No. 6054304
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-913-805A-10

Query Match 95.3% Score 2949; DB 3: Length 575;
Best Local Similarity 95.5%; Pred. No. 1.1e-276;
Matches 549; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 1 MRPTGSRWMLLLFAGTLLFYTGCHLVNDHSDHSSRELKTLAKLERLKOQNDL 60
DB 1 MRPTGSRWMLLLFAGTLLFYTGCHLVNDHSDHSSRELKTLAKLERLKOQNDL 60
QY 61 RRMAGSLRIPGPIIDOGASGRVRLAEQFMKAKEQIENYKQTKNGCGKHLEILRRRIE 120
DB 61 RRMAGSLRIPGPIIDOGASGRVRLAEQFMKAKEQIENYKQTKNGCGKHLEILRRRIE 120
QY 121 NGAKELWFFLOSSELKTKNLKSGNELORHADEFLSDLGHHRSIMTDLYLSOTDAGDWR 180
DB 121 NGAKELWFFLOSSELKTKNLKSGNELORHADEFLSDLGHHRSIMTDLYLSOTDAGDWR 180
QY 181 EKAADLTVELVORITLYLQNPDKCSKAKLVNINKGCGYGCQLHHVVYCFMIAVGTORT 240
DB 181 EKAADLTVELVORITLYLQNPDKCSKAKLVNINKGCGYGCQLHHVVYCFMIAVGTORT 240

DB 181 EKAADLTVELVORITLYLQNPDKCSKAKLVNINKGCGYGCQLHHVVYCFMIAVGTORT 240
QY 241 LALESHNRVATGGEVFRVSECTDRSGSSGTGHHSGEYKDKNNVOVELPIYDSVHPR 300
DB 241 LILESQNRVATGGEVFRVSECTDRSGISTGHHSGEYKDKNNVOVELPIYDSVHPR 300
QY 301 PPLYPLAVPEDLADRLRVHGDPAVMWVSQFVKYLIRPQPLEKEIEEATKLGFKHPVI 360
DB 301 PPLYPLAVPEDLADRLRVHGDPAVMWVSQFVKYLIRPQPLEKEIEEATKLGFKHPVI 360
QY 361 GVHVRFTDKVGAFAFPIEIEYVHVEEDFQLLARMOVDKRRVYLATDPDALLKEAKTK 420
DB 361 GVHVRFTDKVGAFAFPIEIEYVHVEEDFQLLARMOVDKRRVYLATDPDALLKEAKTK 420
QY 421 YPSTEFISDINSISWAGIHNRTENSLRGVLLDHLFSLQADFLVCTFSSQCRAYEIMQ 480
DB 421 YPSTEFISDINSISWAGIHNRTENSLRGVLLDHLFSLQADFLVCTFSSQCRAYEIMQ 480
QY 481 ALHPDASANFRSLDDIYFEGPNAHNOIAIYPHOPRTGELPMEPGDIIGVAGNHMDGYP 540
DB 481 TLHPDASANFRSLDDIYFEGPNAHNOIAIYAHOPRTADELPMEPGDIIIGVAGNHMDGYS 540
QY 541 KGVNRKLGRTGLYPSYKREKIEYVYKYPPTYEADK 575
DB 541 KGVNRKLGRTGLYPSYKREKIEYVYKYPPTYEADK 575

RESULT 4
US-09-442-629-10
Sequence 10, Application US/09442629
Patent No. 6291219
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,629
FILING DATE: 18-NO. 6291219-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796

INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-442-629-10

Query Match 95.3% Score 2949; DB 4: Length 575;
Best Local Similarity 95.5%; Pred. No. 1,1e-276;
Matches 349; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 1 MRPWTSWIMILFLPAMGTLFLYIGSHLVARDNDSDHSSRELSTKLAKLERLKOONEDL 60
1 MRPWTSWIMILFLPAMGTLFLYIGSHLVARDNDSDHSSRELSTKLAKLERLKOONEDL 60
DB 1 MRPWTSWIMILFLPAMGTLFLYIGSHLVARDNDSDHSSRELSTKLAKLERLKOONEDL 60
QY 61 RRAGSLRIPREGIDGPASGRVRALEBPMAKKEQENTKQTKNGPGKDHEILRRRIE 120
61 RRAGSLRIPREGIDGPASGRVRALEBPMAKKEQENTKQTKNGPGKDHEILRRRIE 120
DB 61 RRAGSLRIPREGIDGPASGRVRALEBPMAKKEQENTKQTKNGPGKDHEILRRRIE 120
QY 121 NGAKELMFLQSEKLLKKNLEGNELORHAEFLSDIGHNHRSTMTDLYLTSOTDGAQDMR 180
121 NGAKELMFLQSEKLLKKNLEGNELORHAEFLSDIGHNHRSTMTDLYLTSOTDGAQDMR 180
DB 121 NGAKELMFLQSEKLLKKNLEGNELORHAEFLSDIGHNHRSTMTDLYLTSOTDGAQDMR 180
QY 181 EKEAKDLTELVRRTYLLQNPKDCSKAKKLVNKNKCGYCOLHHVVCFMATAGTORT 240
181 EKEAKDLTELVRRTYLLQNPKDCSKAKKLVNKNKCGYCOLHHVVCFMATAGTORT 240
DB 181 EKEAKDLTELVRRTYLLQNPKDCSKAKKLVNKNKCGYCOLHHVVCFMATAGTORT 240
QY 241 LALESHNRATATGMEVTPRPVSETCTDRSGSTGHWGSGVKKKNQVYELPYDVSVHR 300
241 LALESHNRATATGMEVTPRPVSETCTDRSGSTGHWGSGVKKKNQVYELPYDVSVHR 300
DB 241 LALESHNRATATGMEVTPRPVSETCTDRSGSTGHWGSGVKKKNQVYELPYDVSVHR 300
QY 301 PPLPLAVPEDLDRLVRYVGDPAVMVVSQVYLLRPOPMLKEIEATKLGCFHPVI 360
301 PPLPLAVPEDLDRLVRYVGDPAVMVVSQVYLLRPOPMLKEIEATKLGCFHPVI 360
DB 301 PPLPLAVPEDLDRLVRYVGDPAVMVVSQVYLLRPOPMLKEIEATKLGCFHPVI 360
QY 361 GVHVRRTDKVGAFAHPRIEYTVHEEDFQLLARRMOYDKKRVYLATDPPALLKEAKR 420
361 GVHVRRTDKVGAFAHPRIEYTVHEEDFQLLARRMOYDKKRVYLATDPPALLKEAKR 420
DB 361 GVHVRRTDKVGAFAHPRIEYTVHEEDFQLLARRMOYDKKRVYLATDPPALLKEAKR 420
QY 421 YPVEYFISDSISWSAGLHNRRTENSLRGYILDIHFLSQADFLVCTFSSQVCVAYEIMQ 480
421 YPVEYFISDSISWSAGLHNRRTENSLRGYILDIHFLSQADFLVCTFSSQVCVAYEIMQ 480
DB 421 YPVEYFISDSISWSAGLHNRRTENSLRGYILDIHFLSQADFLVCTFSSQVCVAYEIMQ 480
QY 481 ALHPDASANFRSDDIYFCGPNAHNOIATYPPHQPRTGEPHPEPGDIGVAGNHMDGY 540
481 ALHPDASANFRSDDIYFCGPNAHNOIATYPPHQPRTGEPHPEPGDIGVAGNHMDGY 540
DB 481 ALHPDASANFRSDDIYFCGPNAHNOIATYPPHQPRTGEPHPEPGDIGVAGNHMDGY 540
QY 541 KGVNRRKLGRTGLPSYKVRREKIEYVRYPTYPEADK 575
541 KGVNRRKLGRTGLPSYKVRREKIEYVRYPTYPEADK 575
DB 541 KGVNRRKLGRTGLPSYKVRREKIEYVRYPTYPEADK 575

RESULT 5
US-08-913-805A-3
Sequence 3, Application US/08913805A
Patent No. 6054304
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naotumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-913-805A-3

Query Match 4.2% Score 131; DB 3: Length 26;
Best Local Similarity 100.0%; Pred. No. 4e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 KQTKNGPGKDHEILRRRIENGAKEL 126
1 KQTKNGPGKDHEILRRRIENGAKEL 126
DB 1 KQTKNGPGKDHEILRRRIENGAKEL 126

RESULT 6
US-09-442-629-3
Sequence 3, Application US/09442629
Patent No. 6291219
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naotumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,629
FILING DATE: 18-NO. 6291219-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
APPLICATION NUMBER: PCT/JP97/00171

FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-442-629-3

Query Match 4.2%, Score 131, DB 4, Length 26;
Best Local Similarity 100.0%; Pred. NO. 4e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 KQTKNGPKGDEHLRRRIENGAKEL 126
DB 1 KQTKNGPKGDEHLRRRIENGAKEL 25
|||||

RESULT 7
US-08-913-805A-12
Sequence 12, Application US/08913805A
Patent No. 6054304
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913.805A
FILING DATE: 7 JAN 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:

NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-913-805A-12

Query Match 4.1%, Score 126, DB 3, Length 25;
Best Local Similarity 96.0%; Pred. NO. 1.2e-05;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 352 KLGFKHPVIGVHVRTDKVGAEAF 376
DB 1 KLGFKHPVIGVHVRTDKVGTEAF 25
|||||

RESULT 8
US-09-442-629-12
Sequence 12, Application US/09442629
Patent No. 6291219
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442.629
FILING DATE: 18-NO. 6291219-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/913.805A
FILING DATE: 7 JAN 1998
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-442-629-12

Query Match 4.1%; Score 126; DB 4; Length 25;
Best Local Similarity 96.0%; Pred. No. 1.2e-05;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 352 KLGFKHPVIGVHVRDVKVGAEEAF 376
DB 1 KLGFKHPVIGVHVRDVKVGAEEAF 25

RESULT 9
US-08-968-752B-4
Sequence 4, Application US/08968752B
Patent No. 6043073
GENERAL INFORMATION:
APPLICANT: Frohman, Michael A.
TITLE OF INVENTION: No. 6043073el Phospholipase D Polypeptide and
TITLE OF INVENTION: DNA Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: ONYX Pharmaceuticals, Inc.
STREET: 3031 Research Drive
CITY: Richmond
STATE: California
COUNTRY: USA
ZIP: 94806
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,752B
FILING DATE: 13-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/025,469
FILING DATE: 05-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gloetta, Gregory J.
REGISTRATION NUMBER: 32,028
REFERENCE/DOCKET NUMBER: ONYX2004
TELECOMMUNICATION INFORMATION:
TELEPHONE: 510-222-9700
TELEFAX: 510-222-9758
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1036 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-968-752B-4

Query Match 3.6%; Score 112.5; DB 3; Length 1036;
Best Local Similarity 18.7%; Pred. No. 0.086;
Matches 98; Conservative 68; Mismatches 167; Indels 191; Gaps 24;

QY 114 ILRRRIENGAK-----ELMFLOSELK--LKNLEGN-ELORHAEFLSDL---GH 158
DB 404 ILKRRKQGVRIETIMLYKEVELALGINSEYTKRTLMRLHNITVMRHPDVSTVIMAH 463
QY 159 HERSIMTD--LYVLSQTDGA-GDMREKAKDLTELVRITVLDNPKDCSKARKVCNIN 215
DB 464 HEKLVITIDGVARVGGIDLAYGRMDNE-----HRLTDGVSKRVYSGPSL----- 509
QY 216 KCGCYGCGQLHHVYVCYFMAVGTORTLALBSHNMRYATGGCHETVFRVYSECTCRSSSSTG 275
DB 510 -----GSLPPAMESME----- 521

QY 276 HMSGVKNKNNVQVELPIVDSV-----HPPRYLPLAVPEDL-----ADRLV 317
DB 522 --SLRLKDKNEPVQNLPIQKSIDDVDSKLGKIGKPR-KFSKPSLYQQLRHHLHDADST 578
QY 318 RV-----HGDPVVMVVSQFVKYLIRPQMLEKEIEE-----ATKKL 353
DB 579 SIDSTNTSGSIRSIRLTQVGVGLHGETRFMHGKDYCNFVFDWQDLKPFADFIDRYSTPRM 638
QY 354 GFHPIVIGVHVRDVKVGAEEAFHPIEETVHVEEDFOLLARMQDKRVLAIDDPAL 413
DB 639 PWHDIASAVH---GKAADVARNHFIORW-----NFTYI-----MKSRTLSYPL 681
QY 414 LKEAKTKPSYEFISDNI-----SWSAGLHNYTENSILRGVILDI----- 454
DB 682 LPSQTTAHELRQVPGSVHANVQLLSAADMSAGI-KYHESIHAAYVHENSRIYI 739
QY 455 -----HFLSQADFLVCTFSQCVRAVEITMOA-----LHPDASANRSLDD 495
DB 740 YIENQFISCADDKV-VFNKIGDAIQRLKAHRENOKYRVYVIFLPLGFGDIST--- 795
QY 496 IYFFGGPNAHNOAIPHQPRTEGELPMEPGDIIGVAGHHMOGY 539
DB 796 ---GGNALQALMHPNRYMCRGENSI-LGQLKAELEGQMINY 834

RESULT 10
US-08-630-915A-12
Sequence 12, Application US/08630915A
Patent No. 6309820
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
APPLICANT: HOFEMAN, No. 6309820h
APPLICANT: KAY, Brian K.
APPLICANT: FOMKES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penlie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNTE
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-12

Query Match Similarity 3.6%; Score 112; DB 4; Length 352;
Best Local Similarity 19.2%; Pred.No.0.017; Mismatches 81; Conservative 60; Matches 124; Indels 156; Gaps 18;

OY 178 DMREKEAK-DLT-----ELVGRRIITYLQNPDKDSKAKKLVCNI-----NKGGCYCOL 224
|::||:||::||::||::||::||::||:
Db 32 DFKEMERVDVTYSRAVMELMTKTIEYLQ-PNAPASRAKLSMINTMKSIRGOEKGPGL-QA 89
|::||:||::||
OY 225 HHVVYCMIANGTORTLALESHNMRYATGGMETVPFRPVSETCTDRSGSSFTGHWSGEVKDK 284
|::||:||::||::||::||:
Db 90 EALLAEAMLKFQ-----RELGDGC-----NFSPALGEVGEA 120
|::||:||::||:
OY 285 NVGVVELPIVDSVNHRPPRLPLAVPEDLDRLVRHGDPAWMMVSOFVTKILIRQPMLKK 344
|::||:||::||::||::||:
Db 121 MKELSE--VKDS-----LMDEVKNFIDPDLNLHKDL-----R 152
|::||:||::||::||:
OY 345 ELEEARTKLGFGHPYIGVAHVRFDDKVGAFAAFPHIREY-----TVHEED 389
|::||:||::||::||::||:
Db 153 ETQHHLKLLKGRRLDGFYKAKKROGKITPEDELRQALESFDPSKETAESMFMLEMDIOV 212
|::||:||::||::||::||:
OY 390 FQLLA---RRMOVDKRKYVLATDOPALLE-----AKTKYP----SYEFTSDNIS 433
|::||:||::||::||::||:
Db 213 SOLSLAQAOQLHYHQAAVOIQQOVTAVERIERIQASSQPRREQPKRMISLFATGSTQ 272
|::||:||::||::||::||:
OY 434 WSAGLNHNRTENSLRGVIIDIHFELSADFLVCFSSQCVAVEITMALHPDASANRSRL 493
|::||:||::||::||:
Db 273 PNGLSHGTPKP-REVOMD-----OPCCRALYDL----- 301
|::||:||::||:
OY 494 DDIVYEGGNAHNQIAIVHPORTECEIPMEPDGIIGVANHMHDGYPCGVNRKLQRGTLY 553
|::||:||::||::||::||:
Db 302 -----EPENESELAKGEDDIITLTNOIDENWIEGMLH--QGSCFF 339

OY 554 P 554
+
Db 340 P 340

RESULT 11
PCT-US93-07261-11
Sequence 11, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: John H. C. Blasdale
STREET: One Girilda Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07261
FILING DATE: 19930805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,531
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-822-7398
TELEFAX: 201-822-7039
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1588 amino acids
TYPE: amino acid
TOPOLOGY: linear

```

MOLECULE TYPE: protein
PCT-US93-07261-11

Query Match          3.6%; Score 111; DB 5; Length 1588;
Best Local Similarity 20.7%; Pred. No. 0.24;
Matches 103; Conservative 66; Mismatches 152; Indels 176; Gaps 24;

OY 24 YIGHLVRNDNDSHSSHELKILLAKLERLK-QONEDLRMAGSLRIPDGSPASGR 82
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 131 FIKOH--KDYEERKEKRMRMLIRSLRDRLRETEOLEKIMAOLESAINELKERASRR 187
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 83 VRALEEQRMKKKEDEIENVKKO-----TKNPGCDHEILRIENGAKELMFLOSPLK 136
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 188 PMMKQMQ-RGKMKDEVDEIKRYDDEQAEEKNGT-KDEETIKDK--GDGYEEI--VETKFYG 240
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 137 LKNLEGNELQHNADEFISDLGHHERSIWTDLYULSOTDGAQM----- 179
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 241 MR-----ENALGELEDYEERYEKKRYTYLKE-DGEDLDKVDEEKLEETGTGFREK 288
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 180 -----REKEAKMDITELVORRITYLQNPKDCSKAK-----KLVCINKGCYYC 222
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 289 PPTTRILVKKRRNKKEOKKXEDKEKLLAEDPDDEKKIKLKDSDDKVVVPVNKN----- 343
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 223 QLNHVIVCFMAYGTORTLAESHNMRV-ATGSGMETVPRPSSEFCTDRSGSGSTHMSGEV 281
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 344 -----KSSPEDEKFRAPDKKRTMTFYRLSEL----- 367
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 282 KDKNQVELPIYDSVHPRPYLPLAVPEDLADRLVRYHG-----DP 323
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 368 -----PFIV---PRKDN-ELAVCGSDMS--KYNGKKLLSTPNPKRRRNKLER 411
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 324 AVMWVSQFVKYLIRPOPLEKEIEEATFKLGFKHPVGIVHR----TDKVGAAEAHP- 378
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 412 KMDELHKFKKKVKKYQKLLEREKRENPDGEPALTPEI--HVIKPSDLMDCKEKSACHPF 469
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 379 -----IEEV-TYHVEEDPOL-----LARMQDDKKRVLYLATDDPALK 415
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 470 KYOPTKGAEKEESHVSKDYQLEHPPKLPREYEGKHVSREYQUDH-----EPPTKLP 522
      ::||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY 416 EAKTPYSYEFISDNSI 432
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 523 EYEGKHVSREYQUDNEV 539

RESULT 12
PCT-US93-07261-16
Sequence 16, Application PC/TUS9307261
GENERAL INFORMATION:
TITLE OF INVENTION: PFEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESSEE: John H. C. Blasdale
STREET: One Ciralda Farms
CITY: Madison
STATE: New Jersey
COUNTRY: USA
ZIP: 07940-1000
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07261
FILING DATE: 19930805
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,531
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Blasdale, John H. C.
REGISTRATION NUMBER: 31,895
REFERENCE/DOCKET NUMBER: DX0288K
TELECOMMUNICATION INFORMATION:

```

```

:
: TELEPHONE: 201-822-7398
: TELEFAX: 201-822-7039
: INFORMATION FOR SEQ ID NO: 16:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1663 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: ORIGINAL SOURCE:
: ORGANISM: Plasmodium falciparum
: STRAIN: Malayan Camp
: PCT-US93-07261-16

Query Match      3.6%; Score 111; DB 5; Length 1663;
Best Local Similarity 20.7%; Pred. No. 0.26;
Matches 103; Conservative 66; Mismatches 152; Indels 176; Gaps 24;

QY 24 YIGGLVLRNDSDHSSRELSTKLAKLEK-QQNEDLRRMAGSLRIPGPIIDOGFASGR 82
: 11 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 131 FLKQHL---KDYEEERKKRRMILRLSRDKLREIOLKLNALQESAJNELKERRASRR 187
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 83 VRLAEQFPAKKEQIENYKKO-----TKNGPGKHELLRRRENGAKELMPFLOSELKK 136
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 188 PMWVKMO-RGMKDEVDWKTKYDDEQAEKNGT-KDEEIKDK--GDGYEEL---VETKEYG 240
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 137 LKNLEGNELQRHADFSLDGHHERSIMTDLVYLSQTDGADW-----179
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 241 MR-----ENALGELDEVEERYEKKRYLYKE-DDEGDLKDYEEKLEETGYGFREK 288
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 180 -----REKAKDLTELVRRIITYLQNPDCSKAK-----KLVGINKGGCYGC 222
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 289 FPTTRLVRRKRNKEKKLKEDEKELIAEEDDEKKIKLDSDDKVVVPVKNK----- 343
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 QLHHVYVCFMAYGTRTLALSHNMRY-ATGGMETFRVPRVSETCDRSGSSGHHMSEV 281
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 344 -----KSSFPDKERRAPDKKRTMEYRLSEL-----367
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 KDKNVOVELPIVDVNRPRRYLPLAVPEDLADRLVRVHG-----DP 323
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 368 -----FRIV---PRKDN-ELAVCGDSMDS--KVNGKKLKSTFNPFKRRRNKLKER 411
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 324 AVMWVSQFVKYLIRPQPMLEKEIEEATKKLGFRHPIYGVHVR-----TKVGAEAFAHP- 378
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 412 KMQLHKFRKNKKYOKLLEREKRENDDEPLNTPEI--HVIRPSDLMDGKENGKSGHPE 469
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 379 -----IEEY-TVHVEEDFOL-----LARRQVQDKRYVLTATDDPALLK 415
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 470 KYOPTKGLKEYESHYSKDYOLEHPEPTKLPREYKGVHVSREYQLDH-----EPPTKLP 522
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 416 EAKTKYPSYEFTSDNSI 432
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 523 EYEKGVHSREYQLDNEV 539
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
US-09-107-149-19
: Sequence 19, Application US/09107149
: Patent No. 6274363
: GENERAL INFORMATION:
: APPLICANT: Leung, David W.
: APPLICANT: Tompkins, Christopher K.
: TITLE OF INVENTION: PHOSPHATIDYLCHOLINE PHOSPHOLIPASE D
: FILE REFERENCE: 077319/0144
: CURRENT APPLICATION NUMBER: US/09/107,149
: EARLIER APPLICATION NUMBER: 1998-06-30
: EARLIER FILING DATE: 1996-12-17
: NUMBER OF SEQ ID NOS: 25
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 19
: LENGTH: 971
: TYPE: PRT
: ORGANISM: Homo sapiens
```

```

US-09-107-149-19

Query Match      3.6%; Score 110.5; DB 4; Length 971;
Best Local Similarity 18.0%; Pred. No. 0.12;
Matches 101; Conservative 67; Mismatches 166; Indels 227; Gaps 24;

QY 114 ILRRRENGAK-----ELMFFLOSELKK--LKNLEGN-ELQRHADFSLD---GH 158
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 404 ILKRAQGVRIEIMLYKREVELALGINSEYTKRTLMRLHPNKVHRPDHVSSTVYLMAN 463
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 HERSIMTD--LVYLSQTDGA-GDMREKAKDLTELVRRIITYLQNPDCSKAKKLVGIN 215
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 464 HEKLYIDDSVAFVGIDILAYGRMDNE-----HRLTVGVSKRYTSQPSL----- 509
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 216 KCGVGCOLHHVYVCFMAYGTRTLALSHNMRYATGGMETFRVPRVSETCDRSGSSGTG 275
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 510 -----GSLPRAAMSMK-----521
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 276 HMSGVKKDNVOVELPI-----VDS-----296
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 522 --SLRLKDKNEBVLPIQKSIDVDSKLKIGKPRKFSFLYQLQRHNLHDADSISS 579
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 297 -----VNRPRRYLPLAVPEDLADR-LVR-----VHG 321
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 580 IDSTSSYFNHYRSHNHLHGKLRPHKLFHPSSSEQGLTRPHADTGIRSLQTVGSELHG 639
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 322 DPAVMWVSQFVKYLIRPQPMLEKEIEE-----ATKKLGFRHPIYGVHVRRTDKVGAFAF 376
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 640 ETRFHHGKDYCFNFKRWYQLDKFRADFDIRKSTGRMPHNDIASAVH---CKARDVAR 695
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 377 HPTEEVTVHVEEDFOLLARRMQVDKKRYVLTATDDPALKEAKTKYPSYEFTSDNSI---- 432
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 696 HFIQRM-----NFKRI-----MKSKYRSLSYFLLPKSQTANHELRYGVPSVHANV 742
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 433 -----SMSAGLHNRYPTENSLRGVILDI-----HFLSQADFLVCTESSQVC 472
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 743 QLRSASDAWSAGI--KYHESTIAAYVHVIENSRIYIENQFFISCADDKV-VFNKIGD 799
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 473 RVAVEIMQA-----LHPDASANFRSLDDIYFGSPNANHQIAIYRPHORTE 518
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 800 AIAQRILKAHRENOKRYRVVYLRPLRGEGDIST-----GGNALQALMHFNVRTCR 852
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 519 GEIPMEPGDIIYAGNHMDGY 539
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 853 GENSI-LGQLKAELGNOWINY 872
: : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
US-08-768-147B-2
: Sequence 2, Application US/08768147B
: Patent No. 5859222
: GENERAL INFORMATION:
: APPLICANT: Leung, David W.
: APPLICANT: Tompkins, Christopher
: TITLE OF INVENTION: PHOSPHATIDYLCHOLINE PHOSPHOLIPASE D
: NUMBER OF SEQUENCES: 14
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cell Therapeutics, Inc.
: STREET: 200 Elliott Avenue West, Suite 400
: CITY: Seattle
: STATE: Washington
: COUNTRY: U.S.A.
: ZIP: 98119
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" disk, 1.44mb, double side, high density
: COMPUTER: Hewlett Packard-IBM Compatible
: OPERATING SYSTEM: MS-DOS Version 6
: SOFTWARE: WORD for WINDOWS
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/768,147B
: FILING DATE: 16-Dec-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:00 ; Search time 35 Seconds
(without alignments)
1824.784 Million cell updates/sec

Title: US-09-839-136-2

Perfect score: 3096

Sequence: 1 MRPWGSMRWIMILIFAMGT.....YKVRKIEYKVPYTPPEADK 575

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3088	99.7	575	18	AAW22124
2	2949	95.3	575	18	AAW22125
3	2610	84.3	515	22	AAW22125
4	1732	55.9	339	22	AAW22125
5	1431.5	46.2	22	22	AAW22125
6	1168	37.7	233	22	AAW22125
7	438	14.1	82	22	AAW22125
8	438	14.1	82	22	AAW22125
9	438	14.1	82	22	AAW22125
10	438	14.1	82	22	AAW22125
11	438	14.1	82	22	AAW22125

12	438	14.1	82	22	AAW15003
13	438	14.1	82	22	AAW27440
14	438	14.1	82	22	AAW2730
15	432	14.0	81	22	AAW4855
16	432	14.0	81	22	AAW3789
17	131	4.2	26	18	AAW22126
18	126	4.1	25	18	AAW22131
19	123	4.0	1951	12	AAW8835
20	120.5	3.9	1427	12	AAW10534
21	119	3.8	1456	22	AAW58673
22	119	3.8	1711	22	AAW79819
23	118.5	3.8	1392	20	AAW06999
24	117.5	3.8	2400	22	ABG20278
25	116.5	3.8	2415	22	ABG20279
26	113.5	3.7	1097	22	ABG57814
27	112.5	3.6	1036	19	AAW53280
28	112	3.6	352	17	AAW05388
29	112	3.6	1527	22	AAW01184
30	111	3.6	398	21	AAW31272
31	111	3.6	437	21	AAW6252
32	111	3.6	1588	15	AAW6605
33	111	3.6	1663	15	AAW46608
34	110.5	3.6	971	21	AAW44717
35	110.5	3.6	971	22	AAW85981
36	110.5	3.6	1074	19	AAW53279
37	110.5	3.6	1074	22	AAW85976
38	110.5	3.6	1125	22	ABG18927
39	110	3.6	976	22	AAW6581
40	110	3.6	981	22	AAW36618
41	110	3.6	3248	17	AAW9795
42	109.5	3.5	762	22	AAW57227
43	109	3.5	214	18	AAW09037
44	109	3.5	214	18	AAW12706
45	109	3.5	229	17	AAW05403

ALIGNMENTS

RESULT 1	
AAW22124	AAW22124 standard; Protein: 575 AA.
XX	
XX	AAW22124;
AC	
XX	
DT	05-MAR-1998 (first entry)
XX	
DE	Pig alpha 1-6 fucosyltransferase.
XX	
XX	
KW	Alpha 1-6 fucosyltransferase; enzyme: pig; human: fucose transfer;
KW	guanosine diphosphate; sugar chain synthesis; modification; antibody;
KW	GlcNAc; cancer diagnosis.
XX	
OS	Sus scrofa.
XX	
XX	
PN	W09727303-A1.
PD	
XX	
PD	31-JUL-1997.
XX	
XX	
PF	23-JAN-1997; 97WO-JP00171.
XX	
PR	22-JUL-1996; 96JP-0192260.
PR	24-JAN-1996; 96JP-0010365.
PR	21-JUN-1996; 96JP-0161648.
PR	24-JUN-1996; 96JP-0162813.
PA	(TOYM) TOYO BOSEKI KK.
XX	
PI	Shiba T, Taniguchi N, Uozumi N, Yanagidani S;
XX	
DR	WPI, 1997-393690/36.
DR	
DR	N-PSDB; AAT76573.
XX	

Peptide #1437 enco
Peptide #1477 enco
Peptide #1412 enco
Human brain expres
Peptide #11826 enc
Fragment #1 of pig
Fragment #2 of hum
Human protein SEQ
Human 160KD mediat
Drosophila melanog
Human protein SEQ
Reslin protein seq
Novel human diagno
Novel human diagno
Drosophila melanog
Human phospholipas
Mouse SH3P4-prote
Rat glutamate tran
Arabidopsis thalia
Arabidopsis thalia
Malaria P1EMP3 ep
Plasmodium falcipa
Human phosphatidyl
Human phosphatidyl
Human phospholipas
Human phosphatidyl
Novel human diagno
Human SCP-1 muteln
Staphylococcus aur
Kinetochore protei
Propionibacterium
Osteoclast stimula
Human clone 5 prot

PT Human or pig alpha 1-6 fucosyl:transferase and DNA encoding it - for
PT synthesis and modification of sugar chains and used as an antigen
PT for production of diagnostic antibodies
PS Claim 4; Page 30-34; 61pp; Japanese.
XX
XX
CC AAM22124 and AAM22125 represent the pig and human alpha 1-6
CC fucosyltransferases of the invention, respectively. The enzyme transfers
CC fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc
CC nearest to R in the receptor molecule: (GlcNAc-beta 1-4-GlcNAc-1-6)
CC (GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta 1-4-GlcNAc-1-6 to give
CC (GlcNAc-beta 1-4-GlcNAc-beta 1-6)(GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta
CC 1-4-GlcNAc-beta 1-4-Fucose-alpha 1-6(GlcNAc-R. It has an optimum pH of about
CC 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5
CC hours at 4 degrees C. The optimum working temperature of the
CC alpha 1-6 fucosyltransferase is 30-37 degrees C. A bivalent metal is
CC not required for activity of the enzyme, and the enzyme is not inhibited
CC in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and
CC modification of sugar chains, and as antigen for the production of
CC antibodies recognising the enzyme. The antibodies can be used for the
CC diagnosis of cancer and other diseases.
XX
XX

SO Sequence 575 AA:

Query Match 99.7%; Score 3088; DB 18; Length 575;

Best Local Similarity 99.8%; Pred. No. 5.7e-289;

Matches 574; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MRWTSQMRIMILILFWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60
DB 1 MRWTSQMRIMILILFWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60
OY 61 RRMAGSLRIPEGIDIDGPAAGRALEEOFMKKEQIENYKKTGKNGPKDHEILRRRIE 120
DB 61 RRMAGSLRIPEGIDIDGPAAGRALEEOFMKKEQIENYKKTGKNGPKDHEILRRRIE 120
OY 121 NGAKELWFLQSELEKRLKLNLEGNELORHNADEFLSDLGHNHRSIMTDLYYLSQTDGAGDMR 180
DB 121 NGAKELWFLQSELEKRLKLNLEGNELORHNADEFLSDLGHNHRSIMTDLYYLSQTDGAGDMR 180
OY 181 EKAQKDLTELVRRIYYLQNPCKDSKAKKLVNCINNGCGCOLHHVYCFMAYGTQRT 240
DB 181 EKAQKDLTELVRRIYYLQNPCKDSKAKKLVNCINNGCGCOLHHVYCFMAYGTQRT 240
OY 241 LALESHNMWRATGMEVFRRVSEETCDRSGSGTHMSGCVKKNQVVELPIYDSVHPR 300
DB 241 LALESHNMWRATGMEVFRRVSEETCDRSGSGTHMSGCVKKNQVVELPIYDSVHPR 300
OY 301 PYYLPLAIVEDLADRLVRVHGDPVAVMWVSQFVYLLIRPOFWLEKEIEATKLGFKHPVI 360
DB 301 PYYLPLAIVEDLADRLVRVHGDPVAVMWVSQFVYLLIRPOFWLEKEIEATKLGFKHPVI 360
OY 361 GVAVRRPDKGAAAFHPIEEYTVHVEEDFOLLARRMOVKKRVYLTATDPAALKEAKTK 420
DB 361 GVAVRRPDKGAAAFHPIEEYTVHVEEDFOLLARRMOVKKRVYLTATDPAALKEAKTK 420
OY 421 YPSYEFTSDSISMSAGLHRYTENSILRGVILDIHFLSQADFLVCTESSQVCAVEIMQ 480
DB 421 YPSYEFTSDSISMSAGLHRYTENSILRGVILDIHFLSQADFLVCTESSQVCAVEIMQ 480
OY 481 ALHPDASANFRSLDDIYFSGPNAHNOIATYHPQRTBEGTIPMEPGDITGVAGNHMDGP 540
DB 481 ALHPDASANFRSLDDIYFSGPNAHNOIATYHPQRTBEGTIPMEPGDITGVAGNHMDGP 540
OY 541 KGVNRKLGRTGLPSYKVRKEITVVKYPTYPEADK 575
DB 541 KGVNRKLGRTGLPSYKVRKEITVVKYPTYPEADK 575

RESULT 2
AAM22125
ID AAM22125 standard; Protein: 575 AA.
XX

AC AAM22125;
XX
XX 05-MAR-1998 (first entry)
DE Human alpha 1-6 fucosyltransferase.
XX
XX Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;
KW guanosine diphosphate; sugar chain synthesis; modification; antibody;
KW GlcNAc; cancer diagnosis.
OS Homo sapiens.
PN MO9727303-A1.
XX 31-JUL-1997.
XX 23-JAN-1997; 97WO-jp00171.
XX
XX 22-JUL-1996; 96JP-0192260.
PR 24-JAN-1996; 96JP-0010365.
PR 21-JUN-1996; 96JP-0161648.
PR 24-JUN-1996; 96JP-0162813.
XX
XX (TOYM) TOYO BOSEKI KK.
XX
XX Shiba T, Taniguchi N, Uozumi N, Yanagidani S;
PI WPI: 1997-393690/36.
XX N-PSDB: AAT76574.
DR
XX
XX
PT Human or pig alpha 1-6 fucosyl:transferase and DNA encoding it - for
PT synthesis and modification of sugar chains and used as an antigen
PT for production of diagnostic antibodies
XX
XX
PS Claim 17; Page 39-43; 61pp; Japanese.
XX
XX AAM22124 and AAM22125 represent the pig and human alpha 1-6
XX fucosyltransferases of the invention, respectively. The enzyme transfers
XX fucose from guanosine diphosphate to the 6-hydroxyl group of the GlcNAc
XX nearest to R in the receptor molecule: (GlcNAc-beta 1-4-GlcNAc-1-6)
XX (GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta 1-4-GlcNAc-1-6 to give
XX (GlcNAc-beta 1-4-GlcNAc-beta 1-6)(GlcNAc-beta 1-2-Man-alpha 1-3)-Man-beta
XX 1-4-GlcNAc-beta 1-4-Fucose-alpha 1-6(GlcNAc-R. It has an optimum pH of about
XX 7.0 (pig) or 7.5 (human), and is stable over the pH range 4-10 after 5
XX hours at 4 degrees C. The optimum working temperature of the
XX alpha 1-6 fucosyltransferase is 30-37 degrees C. A bivalent metal is
XX not required for activity of the enzyme, and the enzyme is not inhibited
XX in the presence of 5 mM EDTA. The enzyme is useful in the synthesis and
XX modification of sugar chains, and as antigen for the production of
XX antibodies recognising the enzyme. The antibodies can be used for the
XX diagnosis of cancer and other diseases.
XX

SO Sequence 575 AA:

Query Match 95.3%; Score 2949; DB 18; Length 575;

Best Local Similarity 95.5%; Pred. No. 1.5e-275;

Matches 549; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

OY 1 MRWTSQMRIMILILFWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60
DB 1 MRWTSQMRIMILILFWGTLFLFYIGGHLVRDNDHSDHSSRELSTLAKLERKQONEDL 60
OY 61 RRMAGSLRIPEGIDIDGPAAGRALEEOFMKKEQIENYKKTGKNGPKDHEILRRRIE 120
DB 61 RRMAGSLRIPEGIDIDGPAAGRALEEOFMKKEQIENYKKTGKNGPKDHEILRRRIE 120
OY 121 NGAKELWFLQSELEKRLKLNLEGNELORHNADEFLSDLGHNHRSIMTDLYYLSQTDGAGDMR 180
DB 121 NGAKELWFLQSELEKRLKLNLEGNELORHNADEFLSDLGHNHRSIMTDLYYLSQTDGAGDMR 180
OY 181 EKAQKDLTELVRRIYYLQNPCKDSKAKKLVNCINNGCGCOLHHVYCFMAYGTQRT 240
DB 181 EKAQKDLTELVRRIYYLQNPCKDSKAKKLVNCINNGCGCOLHHVYCFMAYGTQRT 240

QY	241	LALESHNNRYATGCGMETVPEPVSECTDSDSGSTGHMGSCEVKKANQVVELPVDSVHPR	300
Db	241	LILESQNNRYATGCGMETVPEPVSECTDSDSGISTGHMGSCEVKKANQVVELPVDSLHPR	300
QY	301	PEYLPVAVPEDLADRLVRVHGDPVAVMWVSQFVKYLLRPPQWLEKEIEEATKKLGFKHPI	360
Db	301	PEYLPVAVPEDLADRLVRVHGDPVAVMWVSQFVKYLLRPPQWLEKEIEEATKKLGFKHPI	360
QY	361	GVNVRTDKVGEAEAFNPLEEYTVHVEEDPOLLARKMQDKKVVYLTATDPPALLKEAKTK	420
Db	361	GVNVRTDKVGEAEAFNPLEEYTVHVEEHFOLLARMOYDKKVVYLTATDPPALLKEAKTK	420
QY	421	YPSYEIISDNSTSMWSAGLNNRTYENSLRQVLLIDHFLSQADFLVCTFPSSQVCVAVETMQ	480
Db	421	YPNYEFISDNSTSMWSAGLNNRTYENSLRQVLLIDHFLSQADFLVCTFPSSQVCVAVETMQ	480
QY	481	ALHPDASANFSLDDIYFEGGPNANHQIAYHPQRTGEIPEMPEGDITGVAGNHMDGP	540
Db	481	TLHPDASANFSLDDIYFEGGPNANHQIAYAHQPRTADEIPMEPGDITGVAGNHMDGYS	540
QY	541	KGVNKKLGRTGILYPSKYVREKIEFTVYKPPYTPPEADK	575
Db	541	KGVNKKLGRTGILYPSKYVREKIEFTVYKPPYTPPEAK	575
RESULT 3			
ID	AAAG73884	AAAG73884	
XX	AAAG73884	standard; Protein; 515 AA.	
AC	AAAG73884		
DT	03-SEP-2001	(first entry)	
XX			
DE		Human colon cancer antigen protein SEQ ID NO:4648.	
XX			
KW		Human; colon cancer; colon cancer antigen; diagnosis; detection;	
KW		colorectal carcinoma; chromosome 14.	
XX			
OS		Homo sapiens.	
XX			
PN		WO200122920-A2.	
PD			
XX			
PE		28-SEP-2000; 2000WO-US26524.	
XX			
PR		29-SEP-1999; 99US-0157137.	
PR		03-NOV-1999; 99US-0163280.	
XX			
PA		(HUMA-). HUMAN GENOME SCI INC.	
XX			
PI		Ruben SM, Barash SC, Birse CE, Rosen CA;	
XX			
DR		WPI; 2001-235357/24.	
DR		N-PSDB; AAH33315.	
PT			
XX			
PS		Claim 11; Page 6451-6453; 9803pp; English.	
XX			
CC	AAH32943 to AAH37195 and AAH373514 to AAH377788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated Ps,		

CC		by inserting the nucleic acids into a host cell and culturing the cell
CC		to express the proteins. N and P can be used in the prevention, diagnosis
CC		and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC		and AAH77789 represent sequences used in the exemplification of the
CC		present invention.
CC		N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC		missing at time of publication, meaning no sequences are present for
CC		SEQ ID NO:1027 to 1057, 7921 and 7922.
XX		
SQ	Sequence	515 AA:
	Query Match	84.3% Score 2610; DB 22; Length 515;
	Best Local Similarity	95.3%; Pred. No. 7.1e-243;
	Matches	486; Conservative 6; Mismatches 18; Indels 0; Gaps 0
QY	66	SLRIIEGPIDGGPASGRVLALEEFKAKKEOIENNYKKOTKNGPGKHILRRRIENGANE 125
DB	6	SLRIIEGPIDGGPAGRVAVLEEQVLKAKEOEIENYKKOTRNGIGKHIELRRRIENGANE 65
QY	126	LWPFLOSELKKLTKLNLEGNELORNADEFLSDLGHNHERSIIMTDLYLLSOTDAGGWREKAEK 185
DB	66	LWFPELOSELKLTKNLEGNELORNADEFLLDLGHNHERSIIMTDLYLLSOTDAGGWREKAEK 125
QY	186	DLTELVRIRITTLONPKDCSKAKLVICNINKGCGSCOLHHVVYCFMIAVGTOPTALLES 245
DB	126	DLTELVORRITTYLONPKDCSKAKLVICNINKGCGSCOLHHVVYCFMIAVGTOPTALLES 185
QY	246	HNMRVATGMETVFRRPVSETCTDRSGSSGTSHMSGCEVKDKNVGVLEPLVIDSVAPRPYP 305
DB	186	QNMVRATGMETVFRRPVSECTDRSGISGTSHMSGCEVKDKNVGVLEPLVIDSLAPRPYP 245
QY	306	LAVPEDLADRLVHVHGDAVVMVVSQFVKYLIRPOWYLEKEIEBATKKLGFRNPVCVNR 365
DB	246	LAVPEEDLADRLVHVHDPAVVMVVSQFVKYLIRPOWYLEKEIEBATKKLGFRNPVICVNR 305
QY	366	RINDKCAEAAFPRIEEYTVVHEEDPOLLRMVODKKRYLATDDPALLEAKTKPYSE 425
DB	306	RIDKVGTEAFAFPIIEEYVMVHVEHFOLLARRMOVDKKRYYLATADDPSLLEAKTKIPNTE 365
QY	426	FISDNSISMSAGLIHNRYTENSLSRGVILDIHFLSQADFLVCTFSOYCRAVYEIMQLHPD 485
DB	366	FISDNSISMSAGLIHNRYTENSLSRGVILDIHFLSQADFVLCSTFSOYCRAVAELMQTLHPD 425
QY	486	ASANFSLDDIYFYEGGNAHNQIAIYHQPTREGELIMEPCDIIIGVAGNHMDGYPGCVNR 545
DB	426	ASANFSLDDIYFYEGGNAHNQIAIYAHPRTADEIIMEPCDIIIGVAGNHMDGYSKVNR 485
QY	546	KIGRTGLIPSYKYVREKIETVKKPTYPEADK 575
DB	486	KIGRTGLIPSYKYVREKIETVKKPTYPEAREK 515
RESULT 4		
ID	AAAB75061	
AC	AAAB75061	standard; Protein; 339 AA.
XX	AAAB75061;	
DT	20-JUL-2001	(first entry)
XX		
DE	Human alpha 1-6 fucosyltransferase protein 237-575 SEQ ID NO:3.	
XX		
KN	Human: alpha 1-6 fucosyltransferase; alpha 1-6 fucn; antibody;	
XX	alpha 1-6 fucose transferase; anti-human alpha 1-6 fucn; immunossay.	
OS	Homo sapiens.	
PN	JF2001011097-A.	
XX		
PD	16-JAN-2001.	
XX		
Pf	29-JUN-1999;	99JP-0183569.
XX		

PR 29-JUN-1999; 99JP-0183569.
XX
PA (FJRE) FUJIREBIO KK.
XX
DR WPI; 2001-275926/29.
XX N-PSDB; AAE87952.
PT Novel anti-human alpha1-6 fucose transferase antibody useful for
XX immunoassay -
PS Example 2; Page 8-9; 11pp; Japanese.
XX
CC The present invention describes an anti-human alpha 1-6
CC fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
CC transferase) antibody. Also described are: (1) a hybridoma producing
CC the above monoclonal antibody; (2) an immunoassay for detecting human
CC alpha 1-6 fuct by using the above antibody or its antibody fragment;
CC and (3) a reagent used for the above immunoassay. The anti-human
CC alpha 1-6 fuct antibody can be used for immunoassay. The present
CC sequence represents the human alpha 1-6 fuct protein of residues 237
CC to 575, which is used in an example from the present invention.
XX
SQ Sequence 339 AA:
Query Match 55.9%; Score 1732; DB 22; Length 339;
Best Local Similarity 95.0%; Pred. No. 2,3e-158;
Matches 322; Conservative 4; Mismatches 13; Indels 0; Gaps 0;
QY 237 TQRTLAEINMKRYATGCGTETFRPVSSETCTDSSGSGTGMSGEVNDKNQVVELPIVDS 236
DB 1 TORFLIESQNMRYATGCGTETFRPVSSETCTDSSGSGTGMSGEVNDKNQVVELPIVDS 60
QY 297 VHRRPPLPLAIVEDLADRLVHVGDPAVMVSGFVYLLRPQWLEKEIEATKRLGFK 356
DB 61 LHRRPPLPLAIVEDLADRLVHVGDPAVMVSGFVYLLRPQWLEKEIEATKRLGFK 120
QY 357 HPVIGVHVRTDVGVAEAEHPRIEETVHVVEDPQLARRMOYDKKRVYLATDDPALKE 416
DB 121 HPVIGVHVRTDVGTEAAHPRIEETVHVVEEHFQLARRMOYDKKRVYLATDDPSLKE 180
QY 417 AKRKYPEREISNSTSMSGLNHRYTENSLRGVYLDIHFLSQADPLVCFSSQVCHVAY 476
DB 181 AKRKYPEREISNSTSMSGLNHRYTENSLRGVYLDIHFLSQADPLVCFSSQVCHVAY 240
QY 477 EIMQALHPDASANFRSLDDIYFEGCPNANHQIAIYRPHQRTGEIEMEPGDIIGVAGNH 536
DB 241 EIMQALHPDASANFRSLDDIYFEGCPNANHQIAIYAHQRTADEIEMEPGDIIGVAGNH 300
QY 537 DGYRPGVNRKLGRTGLYRPSYKVRKEIETVYKPYTPPEADK 575
DB 301 DGYRPGVNRKLGRTGLYRPSYKVRKEIETVYKPYTPPEAK 339
RESULT 5
ABB59117
ID ABB59117 standard; Protein: 619 AA.
XX
AC ABB59117;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 4143.
XX
KM Drosophila: developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.

XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
PI WPI; 2001-656860/75.
XX N-PSDB; ABL03220.
DR
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
PS Disclosure: SEQ ID NO 4143; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 619 AA:
Query Match 46.2%; Score 1431.5; DB 22; Length 619;
Best Local Similarity 44.4%; Pred. No. 6.4e-129;
Matches 270; Conservative 110; Mismatches 179; Indels 49; Gaps 7;
QY 7 SW-RWMLIFANGTLLEFYIGHLVYRDN-----DHSDBSRKSLKALERKQONEDL 60
DB 14 SWARALIFVLAIGLVYFVVKLTGTGGOAAGESELNARISQALQMLEHTRQNEEL 73
QY 61 RRMAGSLRIPEGRPIDOGPASGRVAL-----EIOFMKAKEQIENYKK 102
DB 74 KQLIDELMSDQ--LDKOSAMKLVQRLLENALNPKLAEVAGRPRESMFESAPRDLGKNN 131
QY 103 QTRNGPCK-----DHETLRRRIENGAKELVFLQSELKRLKNL---EG 142
DB 132 VAGCARPNDLAEVGRDGEFEPSLEYETRRRIQTNGICEINFEPSSELGKRAVAAGNHS 191
QY 143 NELQRADEFLSDLGHNHERSTIMTDLYLISOTDAGDMREKAEKDLTELVQRTTYLQNK 202
DB 192 ADLEESTINQVLLQGADEHKRSLSDMERMRQSDGEYEMVRHKEARDLSDLVQRRRLHNLQNS 251
QY 203 DCSKAKLVGINKKGCGYGCGQLHHVVYCFMIAVGTQRTLAEINMKRYATGCGTETFRPV 262
DB 252 DCOMARKLVCKLNKGGCGYGGQLHHVVYCFVAIATERTLLKRGKRYHKGCHVEYFOY 311
QY 312 SNSCHDAGTANTYNNPCK---PMTQVLYLPIIDSLMPRPYPLAIVEDLAPRLKRLHGD 368
QY 323 PAVWWWSQFVYKYLIRQPMLEKEIEETKLGFKNHYIGVHVRTKVGAAEAHPRIEY 382
DB 369 PAVWWWSQFVYKYLIRQPMLEKEIEETKLGFKNHYIGVHVRTKVGTEAAACHVEEY 428
QY 383 TVHVEEDFOLLARRMQVDKKRVYLATDDPALKEAKTKRPSYFISDNSISMSAGLNHR 442
DB 429 MIVVEDYKRLLEVNGSTVARRIFLASDDAOVIEAKRKIPYOITIDPEYARMASTRY 486
QY 443 TENSRLGVLDIHFLSQADPLVCTFSSQVCHVAYEIMQALHPDASANFRSLDDIYFEGP 502
DB 489 TDTALNGIILDIHLLMSDHLVCTFSSQVCHVAYEIMQTMYPDAANFRSLDDIYFEGG 548
QY 503 NAHNQIAIYRPHQRTGEIEMEPGDIIGVAGNHWDGYPKGVNRKLGRTGLYRPSYKVRKEI 562
||||| : ||||| : : ||||| ||| : : ||||| |||

Db 549 NANNRRVIAHKPRTHEDLQLRVGDVSVAGNHMDGSKGNRTNQGCLFPSPFKEBKV 608
 QY 563 ETVKYPY 570
 : | | | |
 Db 609 DTAKPLPY 616

RESULT 6
 AAB75062
 ID AAB75062 standard; Protein: 233 AA.

XX AAB75062;

XX 20-JUL-2001 (first entry)

XX Human alpha 1-6 fucosyltransferase protein 343-575 SEQ ID NO:4.

XX Human; alpha 1-6 fucosyltransferase; alpha 1-6 fuct; antibody;
 KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fuct; immunoassay.

XX Homo sapiens.

XX JP2001011097-A.

XX 16-JAN-2001.

XX 29-JUN-1999; 99JP-0183569.

XX 29-JUN-1999; 99JP-0183569.

XX (FURE) FUJIREBIO KK.

XX WPI: 2001-275926/29.

XX N-PSDB: AAF87953.

XX Novel anti-human alpha1-6 fucose transferase antibody useful for
 PT immunoassay -

XX Example 2: Page 9-10; 11pp: Japanese.

XX The present invention describes an anti-human alpha 1-6
 CC fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
 CC transferase) antibody. Also described are: (1) a hybridoma producing
 CC the above monoclonal antibody; (2) an immunoassay for detecting human
 CC alpha 1-6 fuct by using the above antibody or its antibody fragment;
 CC and (3) a reagent used for the above immunoassay. The anti-human
 CC alpha 1-6 fuct antibody can be used for immunoassay. The present
 CC sequence represents the human alpha 1-6 fuct protein of residues 343
 CC to 575, which is used in an example from the present invention.

XX Sequence 233 AA:

Query Match 37.7%; Score 1168; DB 22; Length 233;
 Best Local Similarity 94.4%; Pred. No. 3.7e-104;

Matches 220; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 343 EKEIEEATKKGFKHPYGVHVRRTDKVGAEEAFHPIEEYRVHVEDDQLARRMOVDK 402

Db 1 EKEIEEATKKGFKHPYGVHVRRTDKVGAEEAFHPIEEYRVHVEDDQLARRMOVDK 60

QY 403 RYVLAATDDPALKEAKTKYPSYEFISDNSISWSAGLHNRYTENSURGYLIDHFLSQADF 462

Db 61 RYVLAATDDPALKEAKTKYPSYEFISDNSISWSAGLHNRYTENSURGYLIDHFLSQADF 120

QY 463 LVCTFFSSQVCRAVAYEIMQALPQASANFRSLDDIYFEGCPNANHOIAIYHPQRTGEGIP 522

Db 121 LVCTFFSSQVCRAVAYEIMQALPQASANFRSLDDIYFEGCPNANHOIAIYHPQRTGEGIP 180

QY 523 MEPCGIIIVAGNHMDGYPKGVNRKLGRTGLPYSYKVRKEIEYKYPYPEAK 575

Db 181 MEPCGIIIVAGNHMDGYSKGVNRKLGRTGLPYSYKVRKEIEYKYPYPEAK 233

RESULT 7
 ABB28797
 ID ABB28797 standard; Peptide: 82 AA.

XX ABB28797;

XX 01-FEB-2002 (first entry)

XX Peptide #1448 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI: 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID NO 11765; 327bp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting
 CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 82 AA:

Query Match 14.1%; Score 438; DB 22; Length 82;
 Best Local Similarity 98.8%; Pred. No. 2.2e-34;

Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 280 EYKDKNVQVELPIYDSVHPRPPYLPVAPEDLRLVRVHDDPAVMWVSQFVKYLRPQ 339

Db 1 EYKDKNVQVELPIYDSVHPRPPYLPVAPEDLRLVRVHDDPAVMWVSQFVKYLRPQ 60

QY 340 PWLEKEIEEATKKGFKHPYIG 361

XXXXXXXXXXXXXXXXXXXX

Db 61 PWLEKEIEEATKKLGFKHPYIG 82

RESULT 8
ID ABB33985 standard; Peptide; 82 AA.
XX
AC ABB33985;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #1491 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 26620; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX fetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 82 AA:
XX
XX Query Match 14.1%; Score 438; DB 22; Length 82;
XX Best local Similarity 98.8%; Pred. No. 2.2e-34;
XX Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 280 EVKDKNVQVELPIVDSVHRRPYLPLAVPEDLADRLVHGDPRAVWWSQFVKYLIRQ 339
DB 1 EVKDKNVQVELPIVDSLHRRPYLPLAVPEDLADRLVHGDPRAVWWSQFVKYLIRQ 60

QY 340 PWLEKEIEEATKKLGFKHPYIG 361
DB 61 PWLEKEIEEATKKLGFKHPYIG 82

RESULT 9
ID ABB19422 standard; Protein; 82 AA.
XX
AC ABB19422;
XX
XX

DT 23-JAN-2002 (first entry)
XX
DE Protein #1421 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX
PS Claim 15; SEQ ID NO 21192; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX CCA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 82 AA:
XX
XX Query Match 14.1%; Score 438; DB 22; Length 82;
XX Best local Similarity 98.8%; Pred. No. 2.2e-34;
XX Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 280 EVKDKNVQVELPIVDSVHRRPYLPLAVPEDLADRLVHGDPRAVWWSQFVKYLIRQ 339
DB 1 EVKDKNVQVELPIVDSLHRRPYLPLAVPEDLADRLVHGDPRAVWWSQFVKYLIRQ 60

QY 340 PWLEKEIEEATKKLGFKHPYIG 361
DB 61 PWLEKEIEEATKKLGFKHPYIG 82

RESULT 10
ID AAM54747 standard; Protein; 82 AA.
XX
AC AAM54747;
XX
DT 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 26852.
XX

KW Human; brain expressed exon; gene expression analysis; probe:
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
XX Homo sapiens.
XX
XX MO200157275-A2.
XX
XX PD 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX PF 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX P1 Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-483446/52.
XX
XX DR WPI: 2001-483446/52.
XX
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains -
XX
XX PS Example 4; SEQ ID NO: 26352; 650pp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention.
XX
XX SO Sequence 82 AA:
XX
XX Query Match 14.1%; Score 438; DB 22; Length 82;
XX Best Local Similarity 98.8%; Pred. No. 2.2e-34;
XX Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 280 EVKDKNVQVELPIYDSVHPRPYPLAVPEDLADRLVHGDPAVMWVSQFVKYLIRPQ 339
XX DB 1 EVKDKNVQVELPIYDSVHPRPYPLAVPEDLADRLVHGDPAVMWVSQFVKYLIRPQ 60
XX
XX QY 340 PWLEKEIEATKKLGFKHPVIG 361
XX DB 61 PWLEKEIEATKKLGFKHPVIG 82
XX
XX RESULT 11
XX AAM67139
XX ID AAM67139 standard; Protein; 82 AA.
XX
XX AC AAM67139;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 27445.
XX
XX KW Human; bone marrow expressed exon; gene expression analysis; probe:
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX OS Homo sapiens.
XX
XX PM WO200157276-A2.
XX
XX PD 09-AUG-2001.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

PF 30-JAN-2001; 2001WO-US00668.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI: 2001-488900/53.
XX
XX DR WPI: 2001-488900/53.
XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX
XX PS Example 4; SEQ ID NO: 27445; 658pp + Sequence Listing; English.
XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX
XX SQ Sequence 82 AA:
XX
XX Query Match 14.1%; Score 438; DB 22; Length 82;
XX Best Local Similarity 98.8%; Pred. No. 2.2e-34;
XX Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 280 EVKDKNVQVELPIYDSVHPRPYPLAVPEDLADRLVHGDPAVMWVSQFVKYLIRPQ 339
XX DB 1 EVKDKNVQVELPIYDSVHPRPYPLAVPEDLADRLVHGDPAVMWVSQFVKYLIRPQ 60
XX
XX QY 340 PWLEKEIEATKKLGFKHPVIG 361
XX DB 61 PWLEKEIEATKKLGFKHPVIG 82
XX
XX RESULT 12
XX AAM15003
XX ID AAM15003 standard; Protein; 82 AA.
XX
XX AC AAM15003;
XX
XX DT 12-OCT-2001 (first entry)
XX
XX DE Peptide #1437 encoded by probe for measuring cervical gene expression.
XX
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX
XX OS Homo sapiens.
XX
XX PM WO200157278-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US00670.
XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488901/53.
DR
XX
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
XX Claim 27: SEQ ID No 19829; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
CC (SENPs: see A1110068-A128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 82 AA;
Query Match 14.1%; Score 438; DB 22; Length 82;
Best Local Similarity 98.8%; Pred. No. 2.2e-34;
Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 280 EVKDKNVQVELPIVDSVHRRPYLLPLAVPEDLADRLVHNGPRAVWWSQFYKYLIRPQ 339
DB 1 EVKDKNVQVELPIVDSVHRRPYLLPLAVPEDLADRLVHNGPRAVWWSQFYKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFKHPYIG 361
DB 61 PWLEKEIEEATKKLGFKHPYIG 82
DE Peptide #1477 encoded by probe for measuring placental gene expression.
XX
XX Probe: microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200157272-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00663.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-48897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human placenta -
XX
XX Claim 27: SEQ ID No 27709; 654pp; English.
XX
XX
XX The present invention relates to single exon nucleic acid probes (SENPs:
CC see A1131315-A1157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX
XX
SQ Sequence 82 AA;
Query Match 14.1%; Score 438; DB 22; Length 82;
Best Local Similarity 98.8%; Pred. No. 2.2e-34;
Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 280 EVKDKNVQVELPIVDSVHRRPYLLPLAVPEDLADRLVHNGPRAVWWSQFYKYLIRPQ 339
DB 1 EVKDKNVQVELPIVDSVHRRPYLLPLAVPEDLADRLVHNGPRAVWWSQFYKYLIRPQ 60
QY 340 PWLEKEIEEATKKLGFKHPYIG 361
DB 61 PWLEKEIEEATKKLGFKHPYIG 82
DE Peptide #1412 encoded by probe for measuring breast gene expression.
XX
XX Probe: human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00661.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX Claim 27: SEQ ID No 11470; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
CC (see A1100010-A1110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic

CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 82 AA:

Query Match 14.1%; Score 438; DB 22; Length 82;
Best Local Similarity 98.8%; Pred. No. 2.2e-34;
Matches 81; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 280 EVKDKNVQVELPIVDSYHPPRYLPLAVPEDLADRLVHGDPAVMWSQFVKYLIRPQ 339
DB 1 EVKDKNVQVELPIVDSYHPPRYLPLAVPEDLADRLVHGDPAVMWSQFVKYLIRPQ 60
|||||

OY 340 PWLEKEIEATKKGFKHPVIG 361
DB 61 PWLEKEIEATKKGFKHPVIG 82
|||||

RESULT 15
AAM64859

ID AAM64859 standard; Protein: 81 AA.

XX AAM64859;

XX 05-NOV-2001 (first entry)

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 36964.

KW Human: brain expressed exon; gene expression analysis; probe:
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.

XX Homo sapiens.

PN WO200157275-A2.

XX 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

DR Single exon nucleic acid probes for analyzing gene expression in human

XX brains -

XX Example 4; SEQ ID NO: 36964; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX Sequence 81 AA:

Query Match 14.0%; Score 432; DB 22; Length 81;
Best Local Similarity 98.8%; Pred. No. 8.3e-34;
Matches 80; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 280 EVKDKNVQVELPIVDSYHPPRYLPLAVPEDLADRLVHGDPAVMWSQFVKYLIRPQ 339
DB 1 EVKDKNVQVELPIVDSYHPPRYLPLAVPEDLADRLVHGDPAVMWSQFVKYLIRPQ 60
|||||

OY 340 PWLEKEIEATKKGFKHPVI 360
DB 61 PWLEKEIEATKKGFKHPVI 81
|||||

Search completed: October 31, 2002, 13:08:08
Job time : 39 secs

THIS PAGE BLANK (USPTO)

181 AGGAGATGCTGATCTCTCCGAATACCAAGAGGCCCATTTGATGAGGGCCAGCTTCA 240
181 AGGAGATGCTGATCTCTCCGAATACCAAGAGGCCCATTTGATGAGGGCCAGCTTCA 240
241 GGAAGATTCGCTCTTGAAGAGCAATTTATGAAGGCCCAAGACAGATTTGAATTTAT 300
241 GGAAGATTCGCTCTTGAAGAGCAATTTATGAAGGCCCAAGACAGATTTGAATTTAT 300
301 AGAACAACAACTAAAAATGCTCCAGGAAGATCATGAATCTTAAGAGAGAGATTTGA 360
301 AGAACAACAACTAAAAATGCTCCAGGAAGATCATGAATCTTAAGAGAGAGATTTGA 360
361 AATGAGACTTAAGAGCTTCGCTTTTCTTCAAGTGAAGTGAAGAAATTTAAGAAATTTA 420
361 AATGAGACTTAAGAGCTTCGCTTTTCTTCAAGTGAAGTGAAGAAATTTAAGAAATTTA 420
421 GAAGAAATGAACTCCAAAGACATGCAATGATGATTTCTATCAGATTTGGACATGAA 480
421 GAAGAAATGAACTCCAAAGACATGCAATGATGATTTCTATCAGATTTGGACATGAA 480
481 AGCTATATATGAGATCTATCTACTACCTGACCAAAACAGATGGGCGAGTGTGGCGT 540
481 AGCTATATATGAGATCTATCTACTACCTGACCAAAACAGATGGGCGAGTGTGGCGT 540
541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGCGAGAAATTAACATTTCTTCAAGAT 600
541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGCGAGAAATTAACATTTCTTCAAGAT 600
601 CCCAAGGACTGCAGCAAGAGCTAAGTGTATATCAACAAAGGCTGTGGCTAT 660
601 CCCAAGGACTGCAGCAAGAGCTAAGTGTATATCAACAAAGGCTGTGGCTAT 660
661 GGCTGTACGCTCATATGATAGTGTACTGCTTATGATGATGATGATGATGATGATGATGAT 720
661 GGCTGTACGCTCATATGATAGTGTACTGCTTATGATGATGATGATGATGATGATGATGAT 720
721 CTGCGCTTGAATCTCACAATTTGGCGTACGCTTACGCTTACGCTTACGCTTACGCTTACG 780
721 CTGCGCTTGAATCTCACAATTTGGCGTACGCTTACGCTTACGCTTACGCTTACGCTTACG 780
781 CCGTGAAGTGAAGAGCTGCAGACAGATCTGCGAGCTCCACTGAGACATTTGTGACAGTGAA 840
781 CCGTGAAGTGAAGAGCTGCAGACAGATCTGCGAGCTCCACTGAGACATTTGTGACAGTGAA 840
841 GTAAAGGACAAATGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 900
841 GTAAAGGACAAATGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 900
901 CCGTCAATTTTACCCCTGCTGCTGCCAGAAAGCTTGCAGATGACATTTGTGACAGTGAA 960
901 CCGTCAATTTTACCCCTGCTGCTGCCAGAAAGCTTGCAGATGACATTTGTGACAGTGAA 960
961 CCGTCAATTTTACCCCTGCTGCTGCCAGAAAGCTTGCAGATGACATTTGTGACAGTGAA 1020
961 CCGTCAATTTTACCCCTGCTGCTGCCAGAAAGCTTGCAGATGACATTTGTGACAGTGAA 1020
1021 TGCGTGAAGAAAGAAATGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1080
1021 TGCGTGAAGAAAGAAATGTTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 1080
1081 GGAGTCCATTTAGACGACACAGCAAAAGTGGAGCGAAGCAGCTTCCATCCATTGAG 1140
1081 GGAGTCCATTTAGACGACACAGCAAAAGTGGAGCGAAGCAGCTTCCATCCATTGAG 1140
1141 GAATACACGCTGACGTTGAGAAAGATTTGACGCTTCTCTGCGACAGATGCAAGTGAT 1200
1141 GAATACACGCTGACGTTGAGAAAGATTTGACGCTTCTCTGCGACAGATGCAAGTGAT 1200
1201 AAAAAAGGCTGATTTGGCAGACAGATGACCCCTGCTTGTAAAGAGGCAAAAAACAAG 1260
1201 AAAAAAGGCTGATTTGGCAGACAGATGACCCCTGCTTGTAAAGAGGCAAAAAACAAG 1260

1261 TACCCAGATTATGAATTTATTTAGTGAATTAACCTATCTCTTGTGCTGACCTGACATTAAT 1320
1261 TACCCAGATTATGAATTTATTTAGTGAATTAACCTATCTCTTGTGCTGACCTGACATTAAT 1320
1321 CGATATACGAAATTTACCTTGGGGTGTGATCTGATATACCTTCTCTCCAGGCA 1380
1321 CGATATACGAAATTTACCTTGGGGTGTGATCTGATATACCTTCTCTCCAGGCA 1380
1381 GACTTCTAGTGTGATCTTCTTTCATGCAAGGCTGTGAGAGTTCCTTATGAATCATGCA 1440
1381 GACTTCTAGTGTGATCTTCTTTCATGCAAGGCTGTGAGAGTTCCTTATGAATCATGCA 1440
1441 GCGCTGATCTGATGCTGCTGCGAACTTCCTTGTGATGATGATGATGATGATGATGATGAT 1500
1441 GCGCTGATCTGATGCTGCTGCGAACTTCCTTGTGATGATGATGATGATGATGATGATGAT 1500
1501 GGGCCAAATGCCCAACCAAAATTTGCAATTTATCTCACCACCTGCAACTGAAAGGAA 1560
1501 GGGCCAAATGCCCAACCAAAATTTGCAATTTATCTCACCACCTGCAACTGAAAGGAA 1560
1561 ATCCCATGGAACCTGGAGATTTATTTGATGCTGCTGCAATCATGCTGATGCTATCT 1620
1561 ATCCCATGGAACCTGGAGATTTATTTGATGCTGCTGCAATCATGCTGATGCTATCT 1620
1621 AAAGTGTTAAGAGAAATCTGGAAGAGCGGCTATATCTCTCTTCAAAAGTTGAGAG 1680
1621 AAAGTGTTAAGAGAAATCTGGAAGAGCGGCTATATCTCTCTTCAAAAGTTGAGAG 1680
1681 AAGATGAACAACTCAAGTACCCACATATCCGAGGCTGACAAAGTAA 1728
1681 AAGATGAACAACTCAAGTACCCACATATCCGAGGCTGACAAAGTAA 1728

RESULT 2

US-09-839-136-9
Sequence 9, Application US/09839136
Patent No. US2002081694A1
GENERAL INFORMATION:
APPLICANT: Naoyuki TANIGUCHI et al.
TITLE OF INVENTION: ALPHA 1-6 FUCOSYLTRANSFERASE
FILE REFERENCE: 2356-7
CURRENT APPLICATION NUMBER: US/09/839,136
CURRENT FILING DATE: 2001-04-23
PRIOR APPLICATION NUMBER: 09/442,629
PRIOR FILING DATE: 1999-11-18
PRIOR APPLICATION NUMBER: 08/913,805
PRIOR FILING DATE: 1998-01-07
PRIOR APPLICATION NUMBER: PCT/JP97/00171
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 2100
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (198)...(1925)
US-09-839-136-9

Query Match 87.28; Score 1507.2; DB 10; Length 2100;
Best local similarity 92.0%; Pred. No. 0;
Matches 1590; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

1 ATGCGGCATGAGCTGGTGTGGCGTTGGATTATGCTCATTTCTTTGGCTGGGAGCC 60
198 ATGCGGCATGAGCTGGTGTGGCGTTGGATTATGCTCATTTCTTTGGCTGGGAGCC 257
61 TTGCTATTTTACATAGTGTGCTCACTTTGTACGAGATTAATGACCACTCTGCTTAGC 120
258 TTGCTGTATTATATAGTGTGCTCACTTTGTACGAGATTAATGACCACTCTGCTTAGC 317
121 CGAGAACTGTCCAAAGATTTTGGCAAGCTTGAAAGCTTTAAACAACAATAATGAAGCTTG 180

```

|||||
Db 318 CGAGACTGTCGAAGATTCGGCAAGCTTGAACGCTTAAACAGAGAAATGAAGACTTG 377
Oy 181 AGGAGAAATGGCTGATCTCTCCGAATACAGAAAGGCCCTTATATAGGGCCAGCTTCA 240
Db 378 AGCGCAATGGCCGAATCTCTCCGATACGAAGGCCCTATTTGATGAGGGCCAGCTAT 437
Oy 241 GGAAGAGTTCGCTTTAGAGAGCAATTTATGAGGCCCAAGACAGATTGAAATTTAT 300
Db 438 GGAAGAGTACGCTTTTAAAGAGAGCTTTGTTAAGGCCAAGACAGATTGAAATTTAC 437
Oy 301 AAGAAACAATAAATAATGCTCCAGGGAAGATCATGAAATCTTAAGAGAGAGATTGAA 350
Db 498 AAGAAACAACACCAAGAAATGCTGGGGAAGATCATGAAATCTTAAGAGAGAGATTGAA 557
Oy 361 AATGAGCTTAAGAGAGCTCTGCTTTTCTTACAAAGTAGAGTTGAAGAAATTAAGAAATTA 420
Db 558 AATGAGCTTAAGAGAGCTCTGCTTTTCTTACAAAGTAGAGTTGAAGAAATTAAGAAATTA 617
Oy 421 GAAGAAATGAACCTCCAAAGACATGAGATGAATTTCTATCAGATTGGGACATCATGAA 480
Db 618 GAAGAAATGAACCTCCAAAGACATGAGATGAATTTCTTATTTGATTTAGACATCATGAA 677
Oy 481 AGGTCTATATGACGAGATCTATCTACTCTCAGTCAAAACAGATGGGCGAGTGAATGGCT 540
Db 678 AGGTCTATATGACGAGATCTATCTACTCTCAGTCAAAACAGATGGGCGAGTGAATGGCT 737
Oy 541 GAAAGAGGCCCAAGATCTGACAGAGCTGCTCAGCGGAGATTAACATATCTTCAGAAAT 600
Db 738 GAAAGAGGCCCAAGATCTGACAGAGCTGCTCAGCGGAGATTAACATATCTTCAGAAAT 797
Oy 601 CCCAAGAGCTGCGCAAGAGCGAAGCTAGTGTGAATATCAACAAAGGCTGTGCTAT 660
Db 798 CCCAAGAGCTGCGCAAGAGCGAAGCTAGTGTGAATATCAACAAAGGCTGTGCTAT 857
Oy 661 GGCTGTACGCTCCATCATCTAGTGTACTGCTTATGATTGCAATGAGCCAGCGAACA 720
Db 858 GGCTGTACGCTCCATCATCTAGTGTACTGCTTATGATTGCAATGAGCCAGCGAACA 917
Oy 721 CTCGCTTGGAAATCTCAAAATGGCCCTACGCTACGCTGGGAGATGGAAACTGTGTTAGA 780
Db 918 CTCGCTTGGAAATCTCAAAATGGCCCTACGCTACGCTGGGAGATGGAAACTGTGTTAGA 977
Oy 781 CCTGTAGTGAAGCTGTCAGACAGATGCTGCGAGCTCCAGCTGAGATGGAGTGA 840
Db 978 CCTGTAGTGAAGCTGTCAGACAGATGCTGCGAGCTCCAGCTGAGATGGAGTGA 1037
Oy 841 GTTAAAGACAAAATGTTGAGTGTGAGCTCCCATTTGTAGACAGTGTCTCCCTGCT 900
Db 1038 GTTAAAGACAAAATGTTGAGTGTGAGCTCCCATTTGTAGACAGTGTCTCCCTGCT 1097
Oy 901 CCTCCATATTTTACCCCTGGCTGCCAGAAAGACCTTGGAGATGACCTTGTACGATGCAT 960
Db 1098 CCTCCATATTTTACCCCTGGCTGCCAGAAAGACCTTGGAGATGACCTTGTACGATGCAT 1157
Oy 961 GGTGATCTCGACGCTGTGGTGGTATCCAGTTGTGCAAGTGTGATGTCGCCCAACACC 1020
Db 1158 GGTGATCTCGACGCTGTGGTGGTATCCAGTTGTGCAAGTGTGATGTCGCCCAACACC 1217
Oy 1021 TGGCTGAAAAGAGAAATAGAAAGAGGCCACCAAGAGCTAGAGCTTCAAAACATTCAGTTAT 1080
Db 1218 TGGCTGAAAAGAGAAATAGAAAGAGGCCACCAAGAGCTAGAGCTTCAAAACATTCAGTTAT 1277
Oy 1081 GGAAGTCAATGTTAGACGACAGACAAAGTGGAGGAGGAGACGCTTCCATCCCATTTAG 1140
Db 1278 GGAAGTCAATGTTAGACGACAGACAAAGTGGAGGAGGAGACGCTTCCATCCCATTTAG 1337
Oy 1141 GAATACAGGTGACGCTGTAAGAGAACTTACCTTCTGCTGCGAAGTGAAGAGTGAAT 1200
Db 1338 GAATACAGGTGACGCTGTAAGAGAACTTACCTTCTGCTGCGAAGTGAAGAGTGAAT 1397
Oy 1201 AAAAAAGGCTGATTTTGGCCACAGATGACCTTGTGTTAAAGAGCAAAAGCAAG 1260
|||||

```

```

Db 1398 AAAAAAGAGCTGATTTGGCCACAGATGACCTTCTTTATTAAGAGAGCAAAACAAAG 1457
Oy 1261 TACCCAGTTATATTAATTTATTAATGATTAATCTATCTTTGTCAGCTGACATCAAT 1320
Db 1458 TACCCAGTTATATTAATTTATTAATGATTAATCTATCTTTGTCAGCTGACATCAAT 1517
Oy 1321 CGATATACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1380
Db 1518 CGATATACAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1577
Oy 1381 GACTTCCTAGTGTATCTTTTTCATGCGAGTGTGATGAGTGTCTTATGAATATGACAA 1440
Db 1578 GACTTCCTAGTGTATCTTTTTCATGCGAGTGTGATGAGTGTCTTATGAATATGACAA 1637
Oy 1441 GCGCTGCATTCCTATGCTCTGCGAATCTCCGCTTCTTTGATATACATCTATTTTGA 1500
Db 1638 ACATCTACATCTGATGCTCTGCGAATCTCCGCTTCTTTGATATACATCTATTTTGA 1697
Oy 1501 GGCCCAATTCGCCACAAACCAATTTGCTATTCACCAACCTGCAATGGAAGAGAA 1560
Db 1698 GGCCCAATTCGCCACAAACCAATTTGCTATTCACCAACCTGCAATGGAAGAGAA 1757
Oy 1561 ATCCCATGGAACCTGAGATATTTTGTGTGCTGGAATGACATGAGATGCTATCT 1620
Db 1758 ATCCCATGGAACCTGAGATATTTTGTGTGCTGGAATGACATGAGATGCTATCT 1817
Oy 1621 AAGGTGTTTACGAAATCTGGAAGAGCGGCTATATCTCTCTCAAAATTTCCAGAG 1680
Db 1818 AAGGTGTTTACGAAATCTGGAAGAGCGGCTATATCTCTCTCAAAATTTCCAGAG 1877
Oy 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAGTAA 1728
Db 1878 AAGATGAAACAGTCAAGTACCCACATATCTGAGGCTGAGAAATTA 1925

RESULT 3
US-09-864-761-13359
; Sequence 13359, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemolca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

```



```

APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aemica-X-1
CURRENT FILING DATE: US/09/864,761
PRIORITY FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 18181
LENGTH: 248
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO ALI09847.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.6
OTHER INFORMATION: EXPRESSED IN HL600, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 5.1
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.9
OTHER INFORMATION: SWISSPROT HIT: Q9YLZ7, EVALUATE 6.50e+00
OTHER INFORMATION: NT HIT: AF038280.1, EVALUATE 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AM387766.1, EVALUATE 0.00e+00
US-09-864-761-18181

```

Query Match	12.34	Score 213.4	DB 10	Length 248
Best Local Similarity	91.58	Pred. No. 5e-53		
Matches 226	Conservative	0	Mismatches 21	Indels 0
			Gaps 0	
QY	836	GTGAAGTAAAGGACAAANATGTTACAGTCGTTCGACCTCCCATTTGTACACAGTGTTCAC	895	

Db	1	GTAAAGGAAAGAGCAAAAATGTCACAGTGGTCGAGCTTCCACTTTGTAAGCAACTCTTTCATC	60
QY	896	CTGCTCTCCATATTTTACCCCTGGCTGTCCAGAGACCTTGCAGATGCATTTTGAGAG	955
Db	61	CCCCCTCCATATTTTACCTTGGCTGTACAGAAACCTCCACATGCATCTTGTACGAG	120
QY	956	TCCATGGAGATCCGACGATGTGGTGGTATCCAGTTTTCAAAGTACTTGTATGGCCGAC	1015
Db	121	TGCATGGTGAACCTCGCAGTGTGGTGGTGTCTCAGTTTGTCAAAATCTTGAATCCGCCAC	180
QY	1016	AACCTTGCTGGAAAAAGCAATATGAAGAGAGCCACCAAGAAGCTTAGGCTTCAAAACATTCAG	1075
Db	181	AGCCTTGCTAGAAAAAGAAATAGAAAGAGCCACCAAGAGAGCTTGGCTTCAAAACATTCAG	240
QY	1076	TTATTTG 1082	
Db	241	TTATTTG 247	

US-09-864-761-30284	RESULT 6
Sequence 30284, Application US/09864761	
Patent No. US20020048763A1	
GENERAL INFORMATION:	
APPLICANT: Penn, Sharon G.	
APPLICANT: Rank, David R.	
APPLICANT: Hanzel, David K.	
APPLICANT: Chen, Wenshang	
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR	
FILE REFERENCE: Aecomica-X-1	
CURRENT APPLICATION NUMBER: US/09/864,761	
CURRENT FILING DATE: 2001-05-23	
PRIOR APPLICATION NUMBER: US 60/180,312	
PRIOR FILING DATE: 2000-02-04	
PRIOR APPLICATION NUMBER: US 60/207,456	
PRIOR FILING DATE: 2000-05-26	
PRIOR APPLICATION NUMBER: US 09/632,366	
PRIOR FILING DATE: 2000-08-03	
PRIOR APPLICATION NUMBER: GB 24263,6	
PRIOR FILING DATE: 2000-10-04	
PRIOR APPLICATION NUMBER: US 60/236,359	
PRIOR FILING DATE: 2000-09-27	
PRIOR APPLICATION NUMBER: PCT/US01/00666	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00667	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00664	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00669	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00665	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00668	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00663	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00662	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00661	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: PCT/US01/00670	
PRIOR FILING DATE: 2001-01-30	
PRIOR APPLICATION NUMBER: US 60/234,687	
PRIOR FILING DATE: 2000-09-21	
PRIOR APPLICATION NUMBER: US 09/608,408	
PRIOR FILING DATE: 2000-06-30	
PRIOR APPLICATION NUMBER: US 09/774,203	
PRIOR FILING DATE: 2001-01-29	
NUMBER OF SEQ ID NOS: 49117	
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1	
SEQ ID NO 30284	
LENGTH: 248	


```

: SOFTWARE: Patentlin Ver. 2.1
: SEQ ID NO 3
: LENGTH: 417
: TYPE: DNA
: ORGANISM: Plasmodium falciparum
US-09-351-794A-3

Query Match          2.2%; Score 38.2; DB 10; Length 417;
Best Local Similarity 47.7%; Pred. No. 0.2;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Oy 258 AGAAGACGAATTTATGAAAGCCCAAGACAGATTGAAATTTATGAAGAAACAACCTAATAA 317
      ||||| | | | | | ||||| | | | | | ||||| | | | | | ||||| | | | | |
Db 183 AGAAGAGTTAAATAGGATTGAACAGAAATAATTTGAATATGCAATTAAGAAAGAAATAA 242

Oy 318 TCGTCCAGAGCAAGCATCTGAATCTTAAGCAGAGAGATTGAATAATGCGAGCTAAAGAGCT 377
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 TAAAACTTACAAAGACGCAATTAATATATGATATGATTCAGATTGATTAATAATACATAG 302

Oy 378 CTGCTTTTTTCTACAAAGTGAAGTGAAGAAATTTAAGAAATTTAGAAAGAAATGAACCTCA 437
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 303 AGAACAAATTAGAAAAATTCGAAAAAGAAAAATAAACAAATGGATTAAGATGAATAGACA 362

Oy 438 AAGACATCCAGATCAATTTCTATCAGATTGGGACATCTCTGAAGGCTCTATATG 492
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 363 TAAAAATATATAGAGAAAGAAATTAGACAAATGGATCGATGCAATTTATATGATG 417

RESULT 10
: US-09-351-794A-1
: Sequence 1, Application US/09351794A
: Patent No. US20020042382A1
: GENERAL INFORMATION:
: APPLICANT: DUFFY, PATRICK E.
: APPLICANT: OCKENHOUSE, CHRISTIAN F.
: TITLE OF INVENTION: SEQUESTIN
: FILE REFERENCE: 38644-17519
: CURRENT APPLICATION NUMBER: US/09/351,794A
: CURRENT FILING DATE: 1999-07-13
: PRIOR APPLICATION NUMBER: 08/559,896
: PRIOR FILING DATE: 1995-11-17
: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: Patentlin Ver. 2.1
: SEQ ID NO 1
: LENGTH: 1956
: TYPE: DNA
: ORGANISM: Plasmodium falciparum
US-09-351-794A-1

Query Match          2.2%; Score 38.2; DB 10; Length 1956;
Best Local Similarity 47.7%; Pred. No. 0.54;
Matches 112; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Oy 258 AGAAGACGAATTTATGAAGCCCAAGACAGATTGAAATTTATGAAGAAACAACCTAATAA 317
      ||||| | | | | | ||||| | | | | | ||||| | | | | | ||||| | | | | |
Db 306 AGAAGAGTTAAATAGGATTGAACAGAAATAATTTGAATATGCAATTAAGAAAGAAATAA 365

Oy 318 TCGTCCAGAGCAAGCATCTGAATCTTAAGCAGAGAGATTGAATAATGCGAGCTAAAGAGCT 377
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 366 TAAAACTTACAAAGCGCAATTAATATATGATAGTGTCAAGTTGATTAATATCATAG 425

Oy 378 CTGCTTTTTTCTACAAAGTGAAGTGAAGAAATTTAAGAAATTTAGAAAGAAATGAACCTCA 437
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 AGAAGAAATTAGAAAAATTCGAAAAAGAAAAATAAACAAATGGATTAAGATGAATAGACA 485

Oy 438 AAGACATCCAGATCAATTTCTATCAGATTGGGACATCTGAAGGCTCTATATG 492
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 486 TAAAAATATATAGAGAAAGAAATTAGACAAATGGATCGAATTTATATGATG 540

```



```
Query Match          2.0%: Score 34.8; DB 10; Length 376;
Best Local Similarity 49.5%; Pred. No. 1.9;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 260 AAGAGCAATTATGAGGCCAAGACAGATTGAAAATTATAGAAACAACATAAAATG 319
DB 374 AAAAAATATATAAAAAAGAAAAATATAAAAAATATATACAAAATATACAAAATATAT 315
QY 320 GTCCAGGAGGATCATGAATCTCTAAGAGAGCATTTGAAAATGAGCTAAAGACTCT 379
DB 314 AAAAAAATATATATATATATATATATATATATATATATATATATATATATATAT 255
QY 380 GGTATTTTCTACAAAGTGAAGATTGAAGAATTATAGAAAGAAATGAATCCAAA 439
DB 254 AATTAATATATATATATATATATATATATATATATATATATATATATATATAT 195
QY 440 GA 441
DB 194 AA 193

RESULT 15
US-09-759-841-3
; Sequence 3, Application US/09759841
; Patent No. US20010039026A1
; GENERAL INFORMATION:
; APPLICANT: Rickett, Graham A
; APPLICANT: Dobbs, Susan
; APPLICANT: Perros, Manousos
; TITLE OF INVENTION: Assay Method
; FILE REFERENCE: PC10348APME
; CURRENT APPLICATION NUMBER: US/09/759,841
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: GB 0000661.9
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: GB 0000663.5
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: GB 0000659.3
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Human Immunodeficiency virus type 1
US-09-759-841-3

Query Match          2.0%: Score 34.8; DB 10; Length 1512;
Best Local Similarity 50.6%; Pred. No. 4.6;
Matches 84; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 254 CTTTAGACAGCAATTATGAGGCCAAGACAGATTGAAAATTATAGAAACAACATA 313
DB 1013 CTTTAATATAGATATATATATATATATATATATATATATATATATATATATAT 1072
QY 314 AAAAAATGTCAGGAGGAGATCATGAATCTCTAAGAGAGAGATTGAAAATGAGCTAAG 373
DB 1073 AGCATTCTCAGGAGGAGGAGCAGAAATTTGTGACGCACAGTTTAAATTTGAGGGGAAT 1132
QY 374 AGCTGTGTTTCTACAAAGTGAAGTGAAGAATTAAGAATTT 419
DB 1133 TTTTCTACTGTAATTCACACAACACTGTTTAAATGTAATGTAATGTAATGTAATGTA 1178
```

Search completed: November 2, 2002, 00:06:17
Job time : 118 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2002, 21:34:37 : Search time 60 seconds
(without alignments)
8832.291 Million cell updates/sec

Title: US-09-839-136-1

Perfect score: 1728
Sequence: 1 atgcgcacatgactgctgc.....atcccgagcgtcgaagtaa 1728

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCRTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1726.4	99.9	1728	3	US-08-913-805A-1 Sequence 1, Appl1
2	1726.4	99.9	1728	4	US-09-442-629-1 Sequence 1, Appl1
3	1505.6	87.1	2100	3	US-08-913-805A-9 Sequence 9, Appl1
4	1505.6	87.1	2100	4	US-09-442-629-9 Sequence 9, Appl1
5	75.6	4.4	7218	1	US-08-232-463-14 Sequence 14, Appl1
6	38.2	2.2	417	4	US-08-559-896B-3 Sequence 3, Appl1
7	38.2	2.2	1956	4	US-08-559-896B-1 Sequence 1, Appl1
8	37.8	2.2	1448	2	US-08-037-816A-19 Sequence 19, Appl1
9	37.8	2.2	1448	2	US-08-037-816A-23 Sequence 23, Appl1
10	37.8	2.2	1448	2	US-08-530-146-19 Sequence 19, Appl1
11	37.8	2.2	1448	2	US-08-530-146-23 Sequence 23, Appl1
12	37.8	2.2	1532	2	US-08-037-816A-15 Sequence 15, Appl1
13	37.8	2.2	1532	2	US-08-037-816A-27 Sequence 27, Appl1
14	37.8	2.2	1532	2	US-08-530-146-15 Sequence 15, Appl1
15	37.8	2.2	1532	2	US-08-530-146-27 Sequence 27, Appl1
16	36.8	2.1	4527	4	US-08-944-449-8 Sequence 8, Appl1
17	36.8	2.1	4527	4	US-09-353-362-8 Sequence 8, Appl1
18	36.4	2.1	2960	3	US-08-913-842-3 Sequence 3, Appl1
19	36	2.1	1525	1	US-08-609-572-1 Sequence 1, Appl1
20	36	2.1	1525	4	US-08-841-751-1 Sequence 1, Appl1
21	36	2.1	1525	4	US-08-846-340-1 Sequence 1, Appl1
22	36	2.1	1525	4	US-08-846-344-1 Sequence 1, Appl1
23	35.8	2.1	1461	3	US-08-889-841B-7 Sequence 7, Appl1
24	35.2	2.0	162450	4	US-09-345-882-1 Sequence 1, Appl1
25	35	2.0	14176	1	US-08-307-499-14 Sequence 14, Appl1
26	35	2.0	14176	1	US-08-307-499-14 Sequence 14, Appl1
27	35	2.0	14176	4	US-09-299-268-1 Sequence 1, Appl1

28	35	2.0	14176	4	US-09-299-268-14
29	34.6	2.0	675	1	US-08-307-499-55
30	34.6	2.0	675	4	US-09-299-268-55
31	34.6	2.0	1503	3	US-08-889-841B-1
32	34.6	2.0	1503	3	US-08-889-841B-4
33	34.6	2.0	1539	4	US-07-956-483-21
34	34.6	2.0	1539	4	US-08-472-240A-13
35	34.6	2.0	1759	1	US-08-105-483-279
36	34.6	2.0	1759	1	US-08-709-209-279
37	34.6	2.0	1759	1	US-08-458-101-279
38	34.6	2.0	2020	2	US-08-417-210A-135
39	34.6	2.0	2028	2	US-08-417-210A-138
40	34.6	2.0	2060	2	US-08-417-210A-141
41	34.6	2.0	2571	1	US-08-254-358-3
42	34.6	2.0	2571	1	US-08-475-391-3
43	34.6	2.0	2571	2	US-08-709-609-3
44	34.6	2.0	2571	5	PCT-US95-07178-3
45	34.6	2.0	3807	2	US-08-417-210A-78

ALIGNMENTS

RESULT 1
US-08-913-805A-1
Sequence 1, Application US/08913805A
Patent No. 6054304
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YAMAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
City: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetel, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1728 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

Sequence 14, Appl1
Sequence 55, Appl1
Sequence 55, Appl1
Sequence 1, Appl1
Sequence 4, Appl1
Sequence 21, Appl1
Sequence 13, Appl1
Sequence 279, App
Sequence 279, App
Sequence 135, App
Sequence 138, App
Sequence 141, App
Sequence 3, Appl1
Sequence 3, Appl1
Sequence 3, Appl1
Sequence 3, Appl1
Sequence 78, Appl1

TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-913-805A-1

Query Match 99.98; Score 1726.4; DB 3; Length 1728;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 ATCGGCGCATGAGACTGGTGGTGGGATATGCTCATTTCTTTTGGCTGGGGACC 60
   |||||
DB 1 ATGCGGCGCATGAGACTGGTGGTGGGATATGCTCATTTCTTTTGGCTGGGGACC 60
QY 61 TTGCTATTTTACATAGGTGTCTCCTGTTGATGAGATTAAGCAGCTGTATCAGC 120
   |||||
DB 61 TTGCTATTTTACATAGGTGTCTCCTGTTGATGAGATTAAGCAGCTGTATCAGC 120
QY 121 CGAAGACTGTCAGAGATTTTGGCAAGCTGGAACGCTTAAACAACTGAAAGACTTG 180
   |||||
DB 121 CGAAGACTGTCAGAGATTTTGGCAAGCTGGAACGCTTAAACAACTGAAAGACTTG 180
QY 181 AGAGAAATGCTGATCTCTCCGAATACAGAAAGCCCATTTGATCAGGGGCGACCTTCA 240
   |||||
DB 181 AGAGAAATGCTGATCTCTCCGAATACAGAAAGCCCATTTGATCAGGGGCGACCTTCA 240
QY 241 GGAAGAGTTCTGCTTTAGAGAGCAATTTATGAAAGCCAAAGACAGATTGAAATAT 300
   |||||
DB 241 GGAAGAGTTCTGCTTTAGAGAGCAATTTATGAAAGCCAAAGACAGATTGAAATAT 300
QY 301 AAGAAACAACTAAAAATGTCAGAGAGATCATGAATTTCTTAAGAGAGAGATTGAA 360
   |||||
DB 301 AAGAAACAACTAAAAATGTCAGAGAGATCATGAATTTCTTAAGAGAGAGATTGAA 360
QY 361 AATGAGACTAAGAGCTGCTGTTTCTCAAAAGTGAAGTGAAGAAATTA 420
   |||||
DB 361 AATGAGACTAAGAGCTGCTGTTTCTCAAAAGTGAAGTGAAGAAATTA 420
QY 421 GAAGAAATGAACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGAGATCATGAA 480
   |||||
DB 421 GAAGAAATGAACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGAGATCATGAA 480
QY 481 AGGTCTATATGACGATCTATCTACTACCTGATCAAAACAATGGGGAGGTGATGGCGT 540
   |||||
DB 481 AGGTCTATATGACGATCTATCTACTACCTGATCAAAACAATGGGGAGGTGATGGCGT 540
QY 541 GAAAGAGGCGCAAAAGATCTGACAGAGCTGTCAGGAGGAAATTAATCTTTCAGAAAT 600
   |||||
DB 541 GAAAGAGGCGCAAAAGATCTGACAGAGCTGTCAGGAGGAAATTAATCTTTCAGAAAT 600
QY 601 CCCAAGACTGACAGCAAGCAAGAGCTAGTGTATATCAACAAGGCTGTGGCTAT 660
   |||||
DB 601 CCCAAGACTGACAGCAAGCAAGAGCTAGTGTATATCAACAAGGCTGTGGCTAT 660
QY 661 GGGTGTACGCTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
   |||||
DB 661 GGGTGTACGCTCATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 721 CTGCGCTTGAATCTGCAATTTGGCGTACGCTACTGAGGAGATGGAATGATGATGATGAT 780
   |||||
DB 721 CTGCGCTTGAATCTGCAATTTGGCGTACGCTACTGAGGAGATGGAATGATGATGATGAT 780
QY 781 CCTGTAAAGTGAAGAGTGCACAGACAGATCTGAGAGCTCCACTGAGACATTTGGTCAAGTGA 840
   |||||
DB 781 CCTGTAAAGTGAAGAGTGCACAGACAGATCTGAGAGCTCCACTGAGACATTTGGTCAAGTGA 840
QY 841 GTTAAAGGACAAAATGTTCAAGGTGTTGAGCTCCCATTTTACAGATGATGATGATGATGAT 900
   |||||
DB 841 GTTAAAGGACAAAATGTTCAAGGTGTTGAGCTCCCATTTTACAGATGATGATGATGATGAT 900
QY 901 CCTGCATATTTACCCCGTGTGTCAGAGAGCTTGCAGATGATGATGATGATGATGATGAT 960
   |||||
DB 901 CCTGCATATTTACCCCGTGTGTCAGAGAGCTTGCAGATGATGATGATGATGATGATGAT 960
QY 961 GGTGATCTGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
   |||||

```

```

DB 961 GGTGATCTGTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
   |||||
QY 1021 TTGCTGAAAAGGAAATAGAAAGGCGCACAGAAAGCTTAGGCTTCAAACTATTCAGTTATT 1080
   |||||
DB 1021 TTGCTGAAAAGGAAATAGAAAGGCGCACAGAAAGCTTAGGCTTCAAACTATTCAGTTATT 1080
QY 1081 GGAATCCATGTTAGACGACAGACAAAGTGGAGCGGAGACAGCTTCCATCCATTTGAG 1140
   |||||
DB 1081 GGAATCCATGTTAGACGACAGACAAAGTGGAGCGGAGACAGCTTCCATCCATTTGAG 1140
QY 1141 GAATACAGCTGACGCTTGAAGAGACTTTCAGCTTCTGCGAGATGCAAGTGGAT 1200
   |||||
DB 1141 GAATACAGCTGACGCTTGAAGAGACTTTCAGCTTCTGCGAGATGCAAGTGGAT 1200
QY 1201 AAAAAAGGCTATTTGGCCACAGATGACCTGCTTTGTTAAAGGCGCAAAACAAAG 1260
   |||||
DB 1201 AAAAAAGGCTATTTGGCCACAGATGACCTGCTTTGTTAAAGGCGCAAAACAAAG 1260
QY 1261 TACCCAGTTATGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
   |||||
DB 1261 TACCCAGTTATGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 CGATATACAGAAATTTCACTTGGGGGTGATGATGATGATGATGATGATGATGATGAT 1380
   |||||
DB 1321 CGATATACAGAAATTTCACTTGGGGGTGATGATGATGATGATGATGATGATGATGAT 1380
QY 1381 GACTTCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
   |||||
DB 1381 GACTTCTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440
QY 1441 GCGCTGATCTGATGCTGCGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
   |||||
DB 1441 GCGCTGATCTGATGCTGCGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
QY 1501 GCGCCAAATGCCCAACAATTTGCCATTTATCTTCCACCAACCTGAGACTGAAGAGAA 1560
   |||||
DB 1501 GCGCCAAATGCCCAACAATTTGCCATTTATCTTCCACCAACCTGAGACTGAAGAGAA 1560
QY 1561 ATCCCATGAGACCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
   |||||
DB 1561 ATCCCATGAGACCTGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
QY 1621 AAGGTGTTAAAGAACTGGAGAGGAGGCTATATCTCTTCAAAAGTTGAGAG 1680
   |||||
DB 1621 AAGGTGTTAAAGAACTGGAGAGGAGGCTATATCTCTTCAAAAGTTGAGAG 1680
QY 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAAGTAA 1728
   |||||
DB 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAAGTAA 1728

```

RESULT 2

US-09-442-629-1

; Sequence 1, Application US/09442629

; Patent No. 6291219

; GENERAL INFORMATION:

APPLICANT: TANIGUCHI, Naoyuki

UOZUMI, Naofumi

SHIBA, Tetsuo

YANAGIDANI, Shusaku

TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenyon & Kenyon

STREET: 1025 Connecticut Avenue, N.W., Suite 600

STATE: DC

CITY: Washington

COUNTRY: US

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3+ Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,629
FILING DATE: 18-No. 6291219-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
APPLICATION NUMBER: PCT/3P97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetel, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1728 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-442-629-1

Query Match 99.9%; Score 1726.4; DB 4; Length 1728;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCGGCATGAGCTGTTGCGGCTGGATTATGCTCATTTCTTTTGCCTGGGGACC 60
DB 1 ATGCGGCATGAGCTGTTGCGGCTGGATTATGCTCATTTCTTTTGCCTGGGGACC 60
QY 61 TTGCTATTTTACATGAGTGTGCTGCTGTTAGCAGATATGACCACCTGATCTAGC 120
DB 61 TTGCTATTTTACATGAGTGTGCTGCTGTTAGCAGATATGACCACCTGATCTAGC 120
QY 121 CGAGAACTGTCAGAAATTTTGGCAAGCTGGAACGCTTAAACACAAATGAAGACTTG 180
DB 121 CGAGAACTGTCAGAAATTTTGGCAAGCTGGAACGCTTAAACACAAATGAAGACTTG 180
QY 181 AGGAGATGCTGATCTCTCCGAATACGAAAGGCCCATTTGATCAGGGGCCAGCTTCA 240
DB 181 AGGAGATGCTGATCTCTCCGAATACGAAAGGCCCATTTGATCAGGGGCCAGCTTCA 240
QY 241 GGAAGAGTGTGCTTTTACAGCAATTTATGAAGGCCAAGAAAGATTTGAATTTAT 300
DB 241 GGAAGAGTGTGCTTTTACAGCAATTTATGAAGGCCAAGAAAGATTTGAATTTAT 300
QY 301 AAGAAACAACCTAAATAATGCTCCAGGGAAGATCATGAATCTTAAGAGAGAGATTGAA 360
DB 301 AAGAAACAACCTAAATAATGCTCCAGGGAAGATCATGAATCTTAAGAGAGAGATTGAA 360
QY 361 AATGAGCTAAAGAGCTGTGTTTTTCTTACAAAGTGAAGTGAAGAAATTAAGAAATTA 420
DB 361 AATGAGCTAAAGAGCTGTGTTTTTCTTACAAAGTGAAGTGAAGAAATTAAGAAATTA 420
QY 421 GAAGAAATGAAGCTCCAAAGACATGAGATGATTTCTATGATTTGGACATCTATGAA 480
DB 421 GAAGAAATGAAGCTCCAAAGACATGAGATGATTTCTATGATTTGGACATCTATGAA 480
QY 481 AGGTCTATATGAGGATCTATCTACCTCAGTCAACAGATGGGCGAGGTGATGGCGT 540
DB 481 AGGTCTATATGAGGATCTATCTACCTCAGTCAACAGATGGGCGAGGTGATGGCGT 540

DB 481 AGGTCTATATGAGGATCTATCTACCTCAGTCAACAGATGGGCGAGGTGATGGCGT 540
QY 541 GAAAGAGAGCCCAAAATCTGACAGAGCTGTCCACCGGAAATTAACATATCTTCAGAA 600
DB 541 GAAAGAGAGCCCAAAATCTGACAGAGCTGTCCACCGGAAATTAACATATCTTCAGAA 600
QY 601 CCCAAGAGCTGCGCAAGAGCAAGAGCTAGTGTATATATCAACAAAGGCTGTGCTAT 660
DB 601 CCCAAGAGCTGCGCAAGAGCAAGAGCTAGTGTATATATCAACAAAGGCTGTGCTAT 660
QY 661 GCGTGTACCTCCATCATCTAGTGTAGTCTGCTTTATGATTTGCATATGAGCCAGCA 720
DB 661 GCGTGTACCTCCATCATCTAGTGTAGTCTGCTTTATGATTTGCATATGAGCCAGCA 720
QY 721 CTGCGCTTGGATCTCAATTTGGGCTTACGCTACTGCGGGGATGGAAGCTGTGTTAGA 780
DB 721 CTGCGCTTGGATCTCAATTTGGGCTTACGCTACTGCGGGGATGGAAGCTGTGTTAGA 780
QY 781 CCTGTAAGTGAAGCTGACAGAGATGCGGAGCTCCACCTGACATTTGGTCAAGTGA 840
DB 781 CCTGTAAGTGAAGCTGACAGAGATGCGGAGCTCCACCTGACATTTGGTCAAGTGA 840
QY 841 GTAAGAGCAAAATATGTAAGTGTGAGTGTGAGCTCCCATTTGTAGACAGTGTTCCTCG 900
DB 841 GTAAGAGCAAAATATGTAAGTGTGAGTGTGAGCTCCCATTTGTAGACAGTGTTCCTCG 900
QY 901 CCTGCATATTTTACCCCTGCGCTGCCAGAAAGACCTTGCAGATGCACTTGTAGAGTCCAT 960
DB 901 CCTGCATATTTTACCCCTGCGCTGCCAGAAAGACCTTGCAGATGCACTTGTAGAGTCCAT 960
QY 961 GGTGATCTGCACTGT 1020
DB 961 GGTGATCTGCACTGT 1020
QY 1021 TGCGTGGAAAGAGAAATAGAGAGGCGCACCAAGAACTAGGCTTCAACATCCAGTTAT 1080
DB 1021 TGCGTGGAAAGAGAAATAGAGAGGCGCACCAAGAACTAGGCTTCAACATCCAGTTAT 1080
QY 1081 GGAATCCATGTTTACGACGACAGCAAGAGGAGGCGGAAAGACCTTCCATCCATTTAG 1140
DB 1081 GGAATCCATGTTTACGACGACAGCAAGAGGAGGCGGAAAGACCTTCCATCCATTTAG 1140
QY 1141 GAATACAGGCTGACGCTTGAAGAACTTTCACCTTCTGCGCAAGATGAAGTGGAT 1200
DB 1141 GAATACAGGCTGACGCTTGAAGAACTTTCACCTTCTGCGCAAGATGAAGTGGAT 1200
QY 1201 AAAAAAGGATGATTTTGGCCACAGATGACCCCTGTTTAAAGAGGCAAAACAAAG 1260
DB 1201 AAAAAAGGATGATTTTGGCCACAGATGACCCCTGTTTAAAGAGGCAAAACAAAG 1260
QY 1261 TACCCAGTTATGAAATTTATGATGATTAACCTATCTTGTGCTGACCTGACATCAAT 1320
DB 1261 TACCCAGTTATGAAATTTATGATGATTAACCTATCTTGTGCTGACCTGACATCAAT 1320
QY 1321 CGATATACAGAAATTTACATCTGGGGGTGTGATGATGATGATGATGATGATGATGAT 1380
DB 1321 CGATATACAGAAATTTACATCTGGGGGTGTGATGATGATGATGATGATGATGATGAT 1380
QY 1381 GACTTCTAGTGTGATCTTTTTCATGCGAGGCTGTAGAGTGTCTTATGAATCATGCA 1440
DB 1381 GACTTCTAGTGTGATCTTTTTCATGCGAGGCTGTAGAGTGTCTTATGAATCATGCA 1440
QY 1441 GCGCTCATCTGATGCTCTGCGGAACCTTCCGTTCTTGTGATGACATCTACTATTTTGG 1500
DB 1441 GCGCTCATCTGATGCTCTGCGGAACCTTCCGTTCTTGTGATGACATCTACTATTTTGG 1500
QY 1501 GCGCCAAATGCGCCCAACCAAAATTTGCAATTTATCCACCAACCTGCAAGTGAAGAGAA 1560
DB 1501 GCGCCAAATGCGCCCAACCAAAATTTGCAATTTATCCACCAACCTGCAAGTGAAGAGAA 1560
QY 1561 ATCCCATGGAACCTGAGATATATTTGGTGTGCGTGAAGATCAGTGGGATGGCTATCC 1620
DB 1561 ATCCCATGGAACCTGAGATATATTTGGTGTGCGTGAAGATCAGTGGGATGGCTATCC 1620

QY 1621 AAGGTGTTAACAGAAAGTGGGAGAGGCGCTATATCCCTCTACAAAGTTCCAGAG 1680
|||||
Db 1621 AAGGTGTTAACAGAAAGTGGGAGAGGCGCTATATCCCTCTACAAAGTTCCAGAG 1680
QY 1681 AAGATGAAACAGTCAAGTACCCCATATATCCCGAGCTGACAGTAA 1728
|||||
Db 1681 AAGATGAAACAGTCAAGTACCCCATATATCCCGAGCTGACAGTAA 1728
RESULT 3
US-08-913-805A-9
Sequence 9, Application US/08913805A
Patent No. 6054304
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: 0020MT, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetli, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-913-805A-9
Query Match 87.1%; Score 1505.6; DB 3; Length 2100;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1589; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

|||||
Db 258 TTGCTCTTTTATATAGGTGCTCACTTGTACGAGATAATGACCATCTGATCCTAGC 317
QY 121 CGAAGCTGTCGAAGATTTTGGCAAGCTGGAAAGCTTTAAACAAAGAAATGAAGCTTG 180
Db 318 CGAAGCTGTCGAAGATTTTGGCAAGCTGGAAAGCTTTAAACAAAGAAATGAAGCTTG 377
QY 181 AGGAGATGCGTGCATCTCTCCGAATACGAGAGGCCCATTTGATGAGGCCAGCTTCA 240
|||||
Db 378 AGGCAATGCGCGAATCTCTCCGGATACGAGAGGCCCATTTGATGAGGCCAGCTTCA 437
QY 241 GGAAGATTGCTGCTTTAGAGAGCAATTTATGAGGCCCAAGACAGATTGAAATTTAT 300
|||||
Db 438 GGAAGATTGCTGCTTTAGAGAGCAATTTATGAGGCCCAAGACAGATTGAAATTTAT 497
QY 301 AAGAACAACACTAAAATGCTCCAGGAAGCATCAAGAAATCTTAAGAGAGAGATTGAA 360
|||||
Db 498 AAGAACAACAGCCGAATGCTGCGGGAAGATCATGAATCTTGAGGAGAGGATTGAA 557
QY 361 AATGAGCTTAAGAGCTCTGTTTTTCTCAAGTGAAGTGAAGAAATTAAGAAATTTA 420
|||||
Db 558 AATGAGCTTAAGAGCTCTGTTTTTCTCAAGTGAAGTGAAGAAATTAAGAAATTTA 617
QY 421 GAAGAAATGAACTCCAAGACATGCAGATGAATTTCTATCAGATTGGGACATCATGAA 480
|||||
Db 618 GAAGAAATGAACTCCAAGACATGCAGATGAATTTCTTGGATTGAGACATCATGAA 677
QY 481 AGGTCTATATAGAGATCTATCTACCTACGTCACAAAGATGGGCGAGTGATGGCGT 540
|||||
Db 678 AGGTCTATATAGAGATCTATCTACCTACGTCACAAAGATGGGCGAGTGATGGCGG 737
QY 541 GAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGGAATTAATATCTTCAGAAAT 600
|||||
Db 738 GAAAGAGGCGCAAGATCTGACAGAGCTGGTCCAGCGGGAATTAATATCTTCAGAAAT 797
QY 601 CCCAAGACTGACGAAAGCCCAAGAACTAGTGTATATCAACAAGCTTGCTAT 660
|||||
Db 798 CCCAAGACTGACGAAAGCCCAAGAACTAGTGTATATCAACAAGCTTGCTAT 857
QY 661 GGCTGTACGCTCATCATGATGATGTTAGTCTGTTATGTTGATGAGCCAGCCAGGACA 720
|||||
Db 858 GGCTGTACGCTCATCATGATGATGTTAGTCTGTTATGTTGATGAGCCAGCCAGGACA 917
QY 721 CTCGCTTGAATCTCAGATTTGGCGCTACGCTACTGGGGGATGGGAAATCTGTTTGA 780
|||||
Db 918 CTCATCTTGAATCTCAGATTTGGCGCTACTGCTGATGATGGGAGATGATTTAGG 977
QY 781 CCTGTAGTGAAGCTGCACAGACAGATCTGCAAGCTCCACTGACATTTGTCAGTGA 840
|||||
Db 978 CCTGTAGTGAAGCTGCACAGACAGATCTGCAAGCTCCACTGACATTTGTCAGTGA 1037
QY 841 GTAAAGGACAAAATGTTAGTGTTGAGTCCCATTTAGACAGTGTTCATCTCGT 900
|||||
Db 1038 GTAAAGGACAAAATGTTAGTGTTGAGTCCCATTTAGACAGTGTTCATCTCGT 1097
QY 901 CCTCATATTTTACCCTGCTGTCAGAGAACCTTGCAGATGACCTTGTAGAGTCCAT 960
|||||
Db 1098 CCTCATATTTTACCCTGCTGTCAGAGAACCTTGCAGATGACCTTGTAGAGTCCAT 1157
QY 961 GGTGATCTGCAAGTGTGGTATCCCACTTTGTCAAGTACTTGTATGCTCCCAACACC 1020
|||||
Db 1158 GGTGATCTGCAAGTGTGGTATCCCACTTTGTCAAGTACTTGTATGCTCCCAACACC 1217
QY 1021 TTGGTGGAAAGAAATAGAGAGGCGACCAAGAGCTAGGCTTCAAAATCTCAGTTAT 1080
|||||
Db 1218 TTGGTGGAAAGAAATAGAGAGGCGACCAAGAGCTAGGCTTCAAAATCTCAGTTAT 1277
QY 1081 GGAATCATTTAGAGCAGACAGAAAGTGGAGCGGAACAGCTTCCATCCATTTGAG 1140
|||||
Db 1278 GGAATCATTTAGAGCAGACAGAAAGTGGAGCGGAACAGCTTCCATCCATTTGAG 1337
QY 1141 GAATPACAGGTGACGTTGAAGAGACTTTCAGCTTCTGCGACAAATGCAAGTGAAT 1200
|||||


```

Db 1338 GAGTACATGGTGATGTTGAAGAACAATTTTCACCTTCCTGGACGCAAGTAATGCAACTGGAC 1397
Qy 1201 AAAAAAAGGCGTATTTTGGCCACAGATGACCCCTGCTTTGTTAAAAGAGGCAAAAAACAAG 1260
Db 1398 AAAAAAAGAGGTATTTTGGCCACAGATGACCCCTCTTTTATTAAGAGGCAAAAAACAAG 1457
Qy 1261 TACCCCAAGTTATGAATTTATTAAGTAATCTGATCTGCTGGTCAAGCTGACTACATAT 1320
Db 1458 TACCCCAATTAAGATTTATTAAGTAATCTGATCTGCTGGTCAAGCTGACTACATAT 1517
Qy 1321 CGATATACAGAAAAATTCATCTTCGGGGGTGTGATCCTGGATATACACTTCTCTCCAGGCA 1380
Db 1518 CGATATACAGAAAAATTCATCTTCGGGGAGTATCCTGGATATACATTTCTCTCAGGCA 1577
Qy 1381 GACTTCTAGTGTGTAATTTTTCATGCGACAGTCTGTAGAGTTGCTTATGAATAATCATGCA 1440
Db 1578 GACTTCTAGTGTGTAATTTTTCATGCGACAGTCTGTGAGTTGCTTATGAATAATTCGAA 1637
Qy 1441 GCGTGCATCCTGATGGCTCTGCGCAACTCCGTTCTTGGATGACATCTACTATTTTGG 1500
Db 1638 ACATACATCCTGTATGCTCTGGAACCTGCATCTCTTATGATGACATCTACTATTTTGG 1697
Qy 1501 GGCCCAATGGCCCAACCAAAATTTGCATTTATCTCCACCAACCTGCACTGAAAGGAA 1560
Db 1698 GGCCCAATGGCCCAACCAATTTGCCATTTATGCTCACCAACCCGAACTGCAATGAA 1757
Qy 1561 ATCCCATGGAACCTGAGATATTTATTTGTGTGGTGTGAAATCAGTGGATGGCTATCT 1620
Db 1758 ATCCCATGGAACCTGAGATATTCATTTGTGTGGTGTGAAATCATTTGGATGGCTATTTCT 1817
Qy 1621 AAAGGTGTAAACAGAAACCTGGAGAGGAGGCGCTATATCCCTCTACAAAGTTGGAAG 1680
Db 1818 AAAGGTGTAAACAGAAATTTGGAGAGGAGGCGCTATATCCCTCTACAAAGTTGGAAG 1877
Qy 1681 AAAGTAAAGACAGTCAAGTACCCACATATCCCGAGCTGACAGATAA 1728
Db 1878 AAGTAAAGACAGTCAAGTACCCACATATCTGAGGCTGAGAAATTA 1925

RESULT 4
US-09-442-629-9
: Sequence 9, Application US/09442629
: Patent No. 6291219
: GENERAL INFORMATION:
: APPLICANT: TANIGUCHI, Naoyuki
: UOZUMI, Naofumi
: SHIBA, Tetsuo
: YANAGIDANI, Shusaku
: TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Kenyon & Kenyon
: STREET: 1025 Connecticut Avenue, N.W., Suite 600
: CITY: Washington
: STATE: DC
: COUNTRY: US
: ZIP: 20036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3+ Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
: SOFTWARE: WordPerfect 6.1 Windows
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/442,629
: FILING DATE: 18-NO. 6291219-1999
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/913,805A
: FILING DATE: 7 JAN 1998
: APPLICATION NUMBER: PCT/JP97/00171
: FILING DATE: 23 JAN 1997
: APPLICATION NUMBER: JP 199260
: FILING DATE: 22 JUL 1996
: APPLICATION NUMBER: JP 162813
:

```

:	FILING DATE: 24 JUN 1996
:	APPLICATION NUMBER: JP 161648
:	FILING DATE: 21 JUN 1996
:	APPLICATION NUMBER: JP 10365
:	FILING DATE: 24 JAN 1996
:	ATTORNEY/AGENT INFORMATION:
:	NAME: Toifenetti, Judith L.
:	REGISTRATION NUMBER: 39,048
:	REFERENCE/DOCKET NUMBER: 2356/3
:	TELECOMMUNICATION INFORMATION:
:	TELEPHONE: 202-429-1776
:	TELEFAX: 202-429-0796
:	INFORMATION FOR SEQ ID NO: 9:
:	SEQUENCE CHARACTERISTICS:
:	LENGTH: 2100 base pairs
:	TYPE: nucleic acid
:	STRANDEDNESS: double
:	TOPOLOGY: linear
:	MOLECULE TYPE: cDNA
:	SEQUENCE DESCRIPTION: SEQ ID NO: 9:
:	US-09-442-629-9
:	
:	Query Match
:	Best Local Similarity 87.1%; Score 1505.6; DB 4; Length 2100;
:	Matches 189; Conservative 0; Mismatches 139; Indels 0; Gaps 0
QY	1 ATGCGGCCATGTGACTGCTTCGTGGCGTTGGATTATGCCTCATCTTTTTCCTCGGGGACC 60
DB	198 ATGGCGCCATGTGACTGCTTCGTGGCGTTGGATTATGCCTCATCTTTTTCCTCGGGGACC 257
QY	61 TTGCTATTTTACATAGGTGCTCCTTGTTAGCAGATTAATACCACCTGTGATCACTTAGC 120
DB	258 TTGCTATTTTAAATAGGTGCTCCTTGTTAGCAGATAAATGACCATCTGATCACTTAGC 317
QY	121 GGAGAAGTCTCCAAGATTTTGGCAAAGCTGGGAAGCGTTAAACAACAAAATAAGACTTG 180
DB	318 CGAAGACGTCTCCAAGATTTTGGCAAAGCTTGAACGTTAAACAGCAGATTAAGACTTG 377
QY	181 AGGAGATGTGGTGTGATCTCTCCGAATACCAGAAAGGCCCATTTGATCAGGGCCAGCTTCA 240
DB	378 AAGCGAATGTGGCGAATCTCTCCGAATACCAGAAAGGCCCATTTGATCAGGGCCAGCTTAA 437
QY	241 GGAAGAGTTCGTGCTTTAGAGACCAATTTATGAAGGCCCAAGAACAGATTGAAAAATTAT 300
DB	438 GGAAGAGTTCGTGCTTTAGAGAGCAGCTTGTAAAGGCCAAGAACAGATTGAAAAATTAC 497
QY	301 AAGAAACAACATMAAAATGTCTCAGGGAAGATCATGAATCTCTAAGGAGGAGATTGAA 360
DB	498 AAGAAACAACCAAGAAATGTCTCGGGAGAGATCATGAATCTCTGAGGAGGAGATTGAA 557
QY	361 AATGAGCTAAAGACCTCGTGTTTTTTTCACAAAGTGAAGTCTTGAAGAATTAAGAATTAA 420
DB	558 AATGAGCTAAAGACCTCTGTTTTTTCCTACAGAGTGAATTAAGAATAAATAAGAACTTAA 617
QY	421 GAAGAAATGAACCTCCAAGACATGCAGATGAATTTCTACAGATTTTGGACATCATGAA 480
DB	618 GAAGAAATGAACCTCCAAGACATGCAGATGAATTTCTTTTGCATTTAGACATCATGAA 677
QY	481 AGCTCTATTAATACGGATTTATTAATCACTCAATCAACAGATGGGGCAGGTATGGCGT 540
DB	678 AGCTCTATTAATACGGATTTATTAATCACTCAATCAACAGATGGGGCAGGTATGGCGG 737
QY	541 GAAAGGAGGCGCAAAAGATGTGACAGAGCTGGTCCAGCGGAGAAATACATATGTTCAGAT 600
DB	738 GAAAGGAGGCGCAAAAGATGTGACAGAGCTGGTCCAGCGGAGAAATACATATGTTCAGAT 797
QY	601 CCCAAGAGCTGACGAAAGCCAAAGAGCTAGTGTGAATATCAACAAAGGCTGTGGCTAT 660
DB	798 CCCAAGAGCTGACGAAAGCCAAAGAGCTGTGTGAATATCAACAAAGGCTGTGGCTAT 857
QY	661 GGGCTGCACTCATCATGTAGTGTACTGCTTTATGATTTGCATATATGGCACCCAGCGACA 720
DB	858 GGGCTGCACTCATCATGTAGTGTACTGCTTTATGATTTGCATATATGGCACCCAGCGACA 917

ADDRESSEE: Cooper & Dunham
 STREET: 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/037.816A
 FILING DATE: 26-MAR-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 41190/JPM/ALM
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 977-9550
 TELEFAX: (212) 664-0525
 TELEX: 422523 COOPUI
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1448 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1439
 OTHER INFORMATION:
 US-08-037-816A-19

Query Match	2.28	Score	37.8	DB	2	Length	1448
Best Local Similarity	52.98	Pred. No.	0.11				
Matches	81	Conservative	0	Mismatches	72	Indels	0
						Gaps	0

Qy	254	CTTTAGAAAGAGCATTTTATGACGCCAAAGACAGTTTGAAATTTATAGAAACAACCTA	313
Db	944	CTTTAAACAAGATGATTATTAATTAAGAGAACAAATTTGAGAAATAAACAATAGCTTTA	1003
Qy	314	AAATGGTCCAGGAAGATCATGAATTCCTAAGGAGAGGATTTGAAATGAGCTAAAG	373
Db	1004	ATTCACCTCTCAGGAGGGGACCCGCAAAATTTGATATGACACAGTTTATTTGGAGGAGAT	1063
Qy	374	AGCTCTGGTTTTTCTACAAAGTAGCTTGAAG	406
Db	1064	TTTTCTACTGTAATTCACACCAACTTTTAAATA	1096

RESULT 9
 US-08-037-816A-23
 Sequence 23, Application US/08037816A
 Patent No. 5869624
 GENERAL INFORMATION:
 APPLICANT: Hasel, Karl W.
 APPLICANT: Maddon, Paul J.
 TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
 TITLE OF INVENTION: THEREOF, AND THERAPEUTIC AND PROPHYLACTIC USES
 TITLE OF INVENTION: THEREOF
 NUMBER OF SEQUENCES: 29
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Cooper & Dunham
 STREET: 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

```

1  COMPUTER: IBM PC compatible
2  OPERATING SYSTEM: PC-DOS/MS-DOS
3  SOFTWARE: Patentin Release #1.24
4  CURRENT APPLICATION DATA:
5  APPLICATION NUMBER: US/08/037,816A
6  FILING DATE: 26-Mar-1993
7  CLASSIFICATION: 435
8  ATTORNEY/AGENT INFORMATION:
9  NAME: White, John P.
10 REGISTRATION NUMBER: 28,678
11 REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
12 TELECOMMUNICATION INFORMATION:
13 TELEPHONE: (212) 977-9550
14 TELEFAX: (212) 664-0525
15 TELEX: 422523 COOPUI
16 INFORMATION FOR SEQ ID NO: 23:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 1448 base pairs
19 TYPE: nucleic acid
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: DNA (genomic)
23 FEATURE:
24 NAME/KEY: CDS
25 LOCATION: 1..1438
26 OTHER INFORMATION:
27 US-08-037-816A-23

```

Query Match	2.2%	Score 37.8	DB 2	Length 1448
Best Local Similarity	52.9%	Pred. No. 0.11		
Matches 81	Conservative	0	Mismatches 72	Indels 0
				Caps 0

Accession	Sequence	Position
QY	AAATGTGCCAGGAGGATCATGAAATCTCTAGAGAGAGATGTGAAATGAGAGCTTAAG	373
Db	CTTTAAACAGATAGTTATTAATTTAAAGAGAACAAATTTGAGATTTAAACATATAGCTTTTA	10033
QY	CTTTAGAGAGGCATTTTATGATAGCCCAAGAGACAGATTTGAAATTTTAAAGAAACAACCTA	313
Db	CTTTAAACAGATAGTTATTAATTTAAAGAGAACAAATTTGAGATTTAAACATATAGCTTTTA	10033
QY	ATCACTCTCTCAGAGGGGACCCGACCAAAATTTGTATGACACAGTTTAAATTTGTGGAGAGAAAT	10673
Db	ATCACTCTCTCAGAGGGGACCCGACCAAAATTTGTATGACACAGTTTAAATTTGTGGAGAGAAAT	10673
QY	AGCTCTGGTTTTTCTACAAGATGAGCTTGAGA	406
Db	TTTTCTACTGTAATTCACACCAACTTTTAAATA	1096

RESULT 10
US-08-530-146-19
Sequence 19, Application US/08530146
Patent No. 5886163
GENERAL INFORMATION:
APPLICANT: Hasel, Karl W.
APPLICANT: Maddox, Paul J.
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
TITLE OF INVENTION: THEREOF, AND THERAPEUTIC AND PROPHYLACTIC USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,146
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/037,816
FILING DATE: 26-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41190/JPM/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOPUI
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 1448 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1439
OTHER INFORMATION:
US-08-530-146-19

Query Match 2.2%; Score 37.8; DB 2; Length 1448;
Best Local Similarity 52.9%; Pred. No. 0.11;
Matches 81: Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 254 CTTTAAAGAGCAATTATGAAGGCCAAGACAGATTGAAATTTAAGAAACAACATA 313
DB 944 CTTTAAACAGATAGTTATTAATTAAGACAAATTTGAGATAAACAATGCTTTA 1003

QY 314 AAAATGCTCAGGGAAGATCATGAATCTTAAGAGAGAGATTGAAATGGAGCTAAG 373
DB 1004 ATCAGCTCCAGAGAGGAGCCAGAAATGTATATGCACAGTTTAAATGGAGAGAA 1063

QY 374 AGCTCTGTTTTTTCTACAAAGTGAAGTGAAGA 406
DB 1064 TTTTCTACTGTAATTCACACAACTGTTTAATA 1096

RESULT 11
US-08-530-146-23
Sequence 23, Application US/08530146
Patent No. 5886163
GENERAL INFORMATION:
APPLICANT: Hasel, Karl W.
APPLICANT: Maddon, Paul J.
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESS: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,146
FILING DATE: 26-MAR-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/037,816
FILING DATE: 26-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41190/JPM/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOPUI
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1448 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1438
OTHER INFORMATION:
US-08-530-146-23

Query Match 2.2%; Score 37.8; DB 2; Length 1448;
Best Local Similarity 52.9%; Pred. No. 0.11;
Matches 81: Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 254 CTTTAAAGAGCAATTATGAAGGCCAAGACAGATTGAAATTTAAGAAACAACATA 313
DB 944 CTTTAAACAGATAGTTATTAATTAAGACAAATTTGAGATAAACAATGCTTTA 1003

QY 314 AAAATGCTCAGGGAAGATCATGAATCTTAAGAGAGAGATTGAAATGGAGCTAAG 373
DB 1004 ATCAGCTCCAGAGAGGAGCCAGAAATGTATATGCACAGTTTAAATGGAGAGAA 1063

QY 374 AGCTCTGTTTTTTCTACAAAGTGAAGTGAAGA 406
DB 1064 TTTTCTACTGTAATTCACACAACTGTTTAATA 1096

RESULT 12
US-08-037-816A-15
Sequence 15, Application US/08037816A
Patent No. 5869624
GENERAL INFORMATION:
APPLICANT: Hasel, Karl W.
APPLICANT: Maddon, Paul J.
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESS: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/037,816A
FILING DATE: 26-MAR-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41190/JPM/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOPUI
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1532 base pairs

```

:
:      TYPE: nucleic acid
:      STRANDEDNESS: single
:      TOPOLOGY: linear
:      MOLECULE TYPE: DNA (genomic)
:      FEATURE:
:      NAME/KEY: CDS
:      LOCATION: 1..1522
:      OTHER INFORMATION:
:
US-08-037-816A-15

Query Match
Best Local Similarity 52.9%; Score 37.8; DB 2; Length 1532;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 254 CTTTAAAGAGCAATTTATGAGGCCAAGACAGATTGAAATTTAAGAAACAACCTA 313
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1028 CTTTAAACAGATAGTTATTAATAATTAAGAGAACAAATTTGAGATTAACATAGTCTTTA 1087
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 314 AAATGGTCCAGGAGGAGATGATGAATCCCTAAGAGAGAGATTGAAATGAGCTTAAG 373
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1088 ATCAGCTCCTCAGGAGGAGACCAGAAATTTGATGACAGATTTAATTTGTGAGAGAGAAAT 1147
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 374 AGCTCTGCTTTTCTTACAAAGTAGTGAAGA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1148 TTTTCTAGTGTATTCACACACACTGTTTAATA 1180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-08-037-816A-27
: Sequence 27, Application US/08037816A
: Patent No. 5869624
: GENERAL INFORMATION:
: APPLICANT: Hasel, Karl W.
: TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
: TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/037,816A
: FILING DATE: 26-MAR-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 41190/JPM/AJM
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
: TELEX: 422523 COOPUI
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1532 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1522
: OTHER INFORMATION:
:

```

```

US-08-037-816A-27

Query Match
Best Local Similarity 52.9%; Score 37.8; DB 2; Length 1532;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 254 CTTTAAAGAGCAATTTATGAGGCCAAGACAGATTGAAATTTAAGAAACAACCTA 313
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1028 CTTTAAACAGATAGTTATTAATAATTAAGAGAACAAATTTGAGATTAACATAGTCTTTA 1087
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 314 AAATGGTCCAGGAGGAGATGATGAATCCCTAAGAGAGAGATTGAAATGAGCTTAAG 373
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1088 ATCAGCTCCTCAGGAGGAGACCAGAAATTTGATGACAGATTTAATTTGTGAGAGAGAAAT 1147
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 374 AGCTCTGCTTTTCTTACAAAGTAGTGAAGA 406
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1148 TTTTCTAGTGTATTCACACACACTGTTTAATA 1180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-08-530-146-15
: Sequence 15, Application US/08530146
: Patent No. 5886163
: GENERAL INFORMATION:
: APPLICANT: Hasel, Karl W.
: TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
: TITLE OF INVENTION: THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.24
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/530,146
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/037,816
: FILING DATE: 26-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 41190/JPM/AJM
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
: TELEX: 422523 COOPUI
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1532 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1522
: OTHER INFORMATION:
:
US-08-530-146-15

Query Match
Best Local Similarity 52.9%; Score 37.8; DB 2; Length 1532;
Matches 81; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2002, 20:05:17 : Search time 264 Seconds

(without alignments)
14740.353 Million cell updates/sec

Title: US-09-839-136-1

Perfect score: 1728
Sequence: 1 atgcgcgcactgactgctc.....atcccgagcgcacagtaaa 1728

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

- 1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1726.4	99.9	1728	18	AAAT76573
2	1510.4	87.4	1759	24	ABA98809
3	1510.4	87.4	2100	18	AAAT76574
4	1510.4	87.4	3291	23	ABV22468
5	1510.4	87.4	3291	23	ABV28283
6	1499	86.7	3007	22	AAH33315
7	1472.6	85.2	2008	24	ABK70031
8	1467.8	84.9	1728	24	ABK70032
9	878.4	50.8	1017	22	AAH79552

ID	Score	Query Match	Length	DB ID	Description
10	833.4	48.2	979	21	AA63891
11	833.4	48.2	979	21	ABK70036
12	827	47.9	979	21	AA63892
13	827	47.9	979	21	ABK70037
14	606.8	35.1	699	22	AAH7953
15	351.6	20.3	2761	23	ABL04601
16	288.4	16.7	394	20	AAV89049
17	275.2	15.9	503	22	AAK11265
18	275.2	15.9	503	22	AAI42879
19	218.8	12.7	551	22	AAK11037
20	218.8	12.7	551	22	AAI42659
21	215	12.4	4682	23	ABA04600
22	213.4	12.3	248	22	ABA47883
23	213.4	12.3	248	22	ABA65776
24	213.4	12.3	248	22	ABA32861
25	213.4	12.3	248	22	AAK14177
26	213.4	12.3	248	22	AAK25142
27	213.4	12.3	248	22	AAK39911
28	213.4	12.3	248	22	AAI20722
29	213.4	12.3	248	22	AAI45939
30	213.4	12.3	248	22	AAI57183
31	213.4	12.3	248	22	AAI06420
32	213.4	12.3	248	22	ABSI4010
33	174.2	10.1	9196	24	ABK70033
34	139.6	8.1	384	22	ABA42755
35	139.6	8.1	384	22	ABA53183
36	139.6	8.1	384	22	ABA22955
37	139.6	8.1	384	22	AAK01429
38	139.6	8.1	384	22	AAK26877
39	139.6	8.1	384	22	AAI11508
40	139.6	8.1	384	22	AAI32785
41	139.6	8.1	384	22	AAI01422
42	139.6	8.1	384	22	ABS01477
43	115.8	6.7	129	22	AAK23927
44	115.8	6.7	129	22	AAI55934
45	58.4	3.4	60	24	ABN32244

ALIGNMENTS

RESULT 1
AAT76573
AAT76573 standard; cDNA to mRNA; 1728 BP.

ID	Score	Query Match	Length	DB ID	Description
XX	1726.4	99.9	1728	18	AAAT76573
XX	1510.4	87.4	1759	24	ABA98809
XX	1510.4	87.4	2100	18	AAAT76574
XX	1510.4	87.4	3291	23	ABV22468
XX	1510.4	87.4	3291	23	ABV28283
XX	1499	86.7	3007	22	AAH33315
XX	1472.6	85.2	2008	24	ABK70031
XX	1467.8	84.9	1728	24	ABK70032
XX	878.4	50.8	1017	22	AAH79552

05-MAR-1998 (first entry)

Pig alpha 1-6 fucosyltransferase gene.

Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer; guanosine diphosphate; sugar chain synthesis; modification; antibody; GlcNAc; cancer diagnosis; ss.

Sus scrofa.

Key: CDS
Location/Qualifiers
1..1728
/*tag= a

MO9727303-A1.

31-JUL-1997.

23-JAN-1997; 97WO-JP00171.

22-JUL-1996; 96JP-0193260.

24-JAN-1996; 96JP-0010365.

21-JUN-1996; 96JP-0161648.

24-JUN-1996; 96JP-0162813.

(TOYV) TOYO BOSEKI KK.

Chinese hamster FU
Antibody productio
Rat F078 cDNA. Ra
Antibody productio
Human alpha 1-6 fu
Drosophila mejunog
EST clone CH100.
Human brain expres
Probe #11565 used
Human brain expres
Probe #11345 used
Drosophila melanog
Human breast cell
Human foetal liver
Probe #11327 for g
Human brain expres
Human brain expres
Human bone marrow
Probe #10655 for g
Probe #14625 used
Probe #25865 used
Probe #6411 used t
Human genome-deriv
Antibody productio
Human breast cell
Human foetal liver
Probe #1421 for ge
Human bone marrow
Probe #1441 for ge
Probe #1471 used t
Human genome-deriv
Human brain expres
Probe #24620 used
Human spliced tran

XX Shiba T, Taniguchi N, Uozumi N, Yanagidani S;
PI MPI: 1997-393690/36.
DR P-PSDB; AAM22124.
XX
XX Human or pig alpha 1-6 fucosyl:transferase and DNA encoding it - for
PT synthesis and modification of sugar chains and used as an antigen
PT for production of diagnostic antibodies
XX
PS Claim 5: Page 30-34; 61pp: Japanese.
CC AAT76573 and AAT76574 represent the coding sequences for the pig and
CC human alpha 1-6 fucosyltransferases of the invention, respectively. The
CC encoded enzyme transfers fucose from guanosine diphosphate to the
CC 6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:
CC (GlcNAc-beta 1-2-Man-alpha 1-6)(GlcNAc-beta 1-2-Man-alpha 1-3)Man-beta
CC 1-4(GlcNAc-beta 1-4(GlcNAc-R to give (GlcNAc-beta 1-2-Man-alpha 1-6)(GlcNAc-beta
CC 1-2-Man-alpha 1-3)Man-beta 1-4(GlcNAc-beta 1-4(Fucose-alpha 1-6)GlcNAc-R. It has
CC an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the
CC pH range 4-10 after 5 hours at 4 degrees C. The optimum working
CC temperature of the enzyme is 30-37 degrees C. A bivalent metal is not
CC required for activity of the enzyme, and the enzyme is not inhibited in
CC the presence of 5 mM EDTA. The enzyme is useful in the synthesis and
CC modification of sugar chains, and as antigen for the production of
CC antibodies recognising the enzyme. The antibodies can be used for the
CC diagnosis of cancer and other diseases.
XX
XX Sequence 1728 BP; 521 A; 362 C; 419 G; 426 T; 0 other;

Query Match 99.9%; Score 1726.4; DB 18; Length 1728;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGGGCCATGAGCTGGTGGCGTGGATGATGCTCTCTTTTGGCTGGGGACC 60
DB 1 ATCGGGCCATGAGCTGGTGGCGTGGATGATGCTCTCTTTTGGCTGGGGACC 60
QY 61 TTGCTATTTTACATAGAGTGGTGGTGGTACGAGATATGACCACTCTGACTAGC 120
DB 61 TTGCTATTTTACATAGAGTGGTGGTGGTACGAGATATGACCACTCTGACTAGC 120
QY 121 CGAGAACTGTCAGAGATTTTGGCAAGCTGGAGCCTTAAACAAATGAAGACTTG 180
DB 121 CGAGAACTGTCAGAGATTTTGGCAAGCTGGAGCCTTAAACAAATGAAGACTTG 180
QY 181 AGAGAGATGGCTGATCTCTCCGAATACAGAAAGGCCCTTATGACGGGGCAGCTTCA 240
DB 181 AGAGAGATGGCTGATCTCTCCGAATACAGAAAGGCCCTTATGACGGGGCAGCTTCA 240
QY 241 GGAAGAGTTCGCTCTTAGAAGACATTTTATGAAGGCCCAAGACAGATTGAATAT 300
DB 241 GGAAGAGTTCGCTCTTAGAAGACATTTTATGAAGGCCCAAGACAGATTGAATAT 300
QY 301 AAGAAACAACTAAAAATGGTCCAGGAGAGATCATGAATCTTAAGAGAGAGATTGAA 360
DB 301 AAGAAACAACTAAAAATGGTCCAGGAGAGATCATGAATCTTAAGAGAGAGATTGAA 360
QY 361 AATGAGACTAAAGAGCTCTGTTTTTTCTCAAAAGTGAAGTGAAGAAATTA 420
DB 361 AATGAGACTAAAGAGCTCTGTTTTTTCTCAAAAGTGAAGTGAAGAAATTA 420
QY 421 GAAGGAATGAACCTCCAAAGACATGAGATGATTTCTATCAGATTTGGACATGAA 480
DB 421 GAAGGAATGAACCTCCAAAGACATGAGATGATTTCTATCAGATTTGGACATGAA 480
QY 481 AGGTCTAATAAGAGATCTATCTACTACCTGAGTCAAAACAGATGGGAGTGAATTGGCGT 540
DB 481 AGGTCTAATAAGAGATCTATCTACTACCTGAGTCAAAACAGATGGGAGTGAATTGGCGT 540
QY 541 GAAAGAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATTAACATTTTCAGAAAT 600
DB 541 GAAAGAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGAGATTAACATTTTCAGAAAT 600

QY 601 CCCAAGACTGACGAAAGCCCAAGAGCTAGTGTAAATATCAACAAGAGCTGTGGCTAT 660
DB 601 CCCAAGACTGACGAAAGCCCAAGAGCTAGTGTAAATATCAACAAGAGCTGTGGCTAT 660
QY 661 GGCTGTACGCTCATCATGTAGTGTACTGCTTTATGATTTGCATATGCGACCCAGGACAA 720
DB 661 GGCTGTACGCTCATCATGTAGTGTACTGCTTTATGATTTGCATATGCGACCCAGGACAA 720
QY 721 CTGCGCTTGGAAATCTCACAATTTGGCGCTACGCTACGCTGGGAGATGGGAACCTGTGTTAGA 780
DB 721 CTGCGCTTGGAAATCTCACAATTTGGCGCTACGCTACGCTGGGAGATGGGAACCTGTGTTAGA 780
QY 781 CCGTGAAGTGAAGAGCTGACAGACAGACATGCGAGCTCCACTGGAGATTGGTCAAGTGA 840
DB 781 CCGTGAAGTGAAGAGCTGACAGACAGACATGCGAGCTCCACTGGAGATTGGTCAAGTGA 840
QY 841 GTAAAGGACAAAATGTTCAGTGGTGTGAGCTCCCAATGTAGACAGTGTTCATCTCGT 900
DB 841 GTAAAGGACAAAATGTTCAGTGGTGTGAGCTCCCAATGTAGACAGTGTTCATCTCGT 900
QY 901 CCTCCATATTTACCCCGGCTGTCCAGAAAGACTTGGACAGATGACCTGTGAGAGTCCAT 960
DB 901 CCTCCATATTTACCCCGGCTGTCCAGAAAGACTTGGACAGATGACCTGTGAGAGTCCAT 960
QY 961 GGTGATCTGCTCAGTGTGTGGGTATCCAGCTTGTCAAGTACTTGTATTCGCCCAACCC 1020
DB 961 GGTGATCTGCTCAGTGTGTGGGTATCCAGCTTGTCAAGTACTTGTATTCGCCCAACCC 1020
QY 1021 TGGCTGGAAGAAAGAAATAGAAAGAGGCCACCAAGCTAGGCTTCAAACTCCAGTTAT 1080
DB 1021 TGGCTGGAAGAAAGAAATAGAAAGAGGCCACCAAGCTAGGCTTCAAACTCCAGTTAT 1080
QY 1081 GAGATCCATGTGTAGACGACAGACAAAGTGGAGCGGAACAGCTTCCATCCATTGAG 1140
DB 1081 GAGATCCATGTGTAGACGACAGACAAAGTGGAGCGGAACAGCTTCCATCCATTGAG 1140
QY 1141 GAATACACGGTGCAGCTTGAAGAAAGCTTTCAGCTTCTGCTGAGAAATGCAAGTGGAT 1200
DB 1141 GAATACACGGTGCAGCTTGAAGAAAGCTTTCAGCTTCTGCTGAGAAATGCAAGTGGAT 1200
QY 1201 AAAAAAGGCTGATTTTGGCCACAGATGACCTGCTTTGTTAAAGAGCAAAACAAAG 1260
DB 1201 AAAAAAGGCTGATTTTGGCCACAGATGACCTGCTTTGTTAAAGAGCAAAACAAAG 1260
QY 1261 TACCCAGTATGAATTTATTAGTGAATACCTATCTCTTGGTGCAGCTGACATCAAT 1320
DB 1261 TACCCAGTATGAATTTATTAGTGAATACCTATCTCTTGGTGCAGCTGACATCAAT 1320
QY 1321 CGATATACAGAAATTCACCTTGGGGTGTGATCCTGGATATACACTTTCCTCCAGGCA 1380
DB 1321 CGATATACAGAAATTCACCTTGGGGTGTGATCCTGGATATACACTTTCCTCCAGGCA 1380
QY 1381 GACTTCTACTGTGACTTTTTCATGCGAGCTGTGAGAGTGTCTTATGAATCATGCAA 1440
DB 1381 GACTTCTACTGTGACTTTTTCATGCGAGCTGTGAGAGTGTCTTATGAATCATGCAA 1440
QY 1441 GCGCTCATCTGATGCTGTGCAACTTCCGTTCTTTGGATGACATCTATATTTTGA 1500
DB 1441 GCGCTCATCTGATGCTGTGCAACTTCCGTTCTTTGGATGACATCTATATTTTGA 1500
QY 1501 GGCCCAATGCCCAACCAAAATTTGCCATTATCTCACCAACCTGAACTGAAGAGAA 1560
DB 1501 GGCCCAATGCCCAACCAAAATTTGCCATTATCTCACCAACCTGAACTGAAGAGAA 1560
QY 1561 ATCCCATGAGACCTGAGATATTATTGGTGTGCTGGAATACGTGGATGCTATCT 1620
DB 1561 ATCCCATGAGACCTGAGATATTATTGGTGTGCTGGAATACGTGGATGCTATCT 1620
QY 1621 AAGGTGTTAAGCAAGAACTGGGAAGAGCGGCTATATCCCTCTCAACAAGTTGAGAG 1680
DB 1621 AAGGTGTTAAGCAAGAACTGGGAAGAGCGGCTATATCCCTCTCAACAAGTTGAGAG 1680

Accession	Gene	Species	Protein	Enzyme	EC	Function	Location	Qualifiers
Db 1277	TACCCCAATTATGATTTATTTAGTGATTA	CT	AACTCTGATTTCTGGTCA	AGCTGACAT	1336			
Oy 1321	CGATATACAGAAAAATTCAC	TT	CGGGGTGGATC	CGTGATATAC	TTTCTCTCCAGCA	1380		
Db 1337	CGATATACAGAAAAATTCAC	TT	CGGGGTGGATC	CGTGATATAC	TTTCTCTCCAGCA	1380		
Oy 1381	GACTTTCTAGTGTGTACT	TTTTT	CATCGCAGAGTCTG	TAGAGTTGCTTATGA	ATCATGTGCA	1440		
Db 1397	GACTTTCTAGTGTGTACT	TTTTT	CATCGCAGAGTCTG	TAGAGTTGCTTATGA	ATCATGTGCA	1440		
Oy 1441	GCCTGTCATCTGATGCTG	CGCCTGCGGAACTTC	CGCTTTTGATGACAT	CTACTATTTTGA	1500			
Db 1457	ACACTACATCTCTGATGCT	CTCTCGCAAACTTC	GCATTTCTTATGATGACAT	CTACTATTTTGA	1516			
Oy 1501	GGCCCAATGCGCCACAC	CAAAATTTGCCATTTAT	CTCTCACCAACCTCGA	ACTGAAAGAGAA	1560			
Db 1517	GGCCCAATGCGCCACAC	CAAAATTTGCCATTTAT	CTCTCACCAACCTCGA	ACTGAAAGAGAA	1560			
Oy 1561	ATCCCATGGAACCTGGA	AGATATATTTGGTGTG	GTGCTGTGGAATAC	ACTGGATGGCTATCT	1620			
Db 1577	ATTCCTCATGGAACCTG	GAGATATATCTTTGGTGTG	GTGCTGTGGAATAC	ACTGGATGGCTATCT	1636			
Oy 1621	AAAGGTGTTAAACAGAA	ACACGAGGAGAGCGG	CTATATCCCTCCAC	AAAGTTGAGAG	1680			
Db 1637	AAAGGTGTTAAACAGAA	ACACGAGGAGAGCGG	CTATATCCCTCCAC	AAAGTTGAGAG	1680			
Oy 1681	AAAGTATGAAACAGTCA	AGTACCCACATATAT	CCCGAGGCTGAC	CAAGTAA	1728			
Db 1697	AAAGTATGAAACAGTCA	AGTACCCACATATAT	CCCGAGGCTGAC	CAAGTAA	1744			
RESULT 3	AAAT76574	standard; cdna to mRNA; 2100 BP.						
AC	AAAT76574:							
DT	05-MAR-1998	(first entry)						
DE	Human alpha 1-6 fucosyltransferase gene.							
KW	Alpha 1-6 fucosyltransferase; enzyme; pig; human; fucose transfer;							
KW	glucosamine diphosphate; sugar chain synthesis; modification; antibody;							
KW	glucNAc; cancer diagnosis; ss.							
OS	Homo sapiens.							
FT	Key	Location/Qualifiers						
FT	CDS	198..1925						
FT		/*tag= a						
PN	MO9727303-A1.							
PE	23-JAN-1997;	97MO-JP00171.						
PR	22-JUL-1996;	96JP-0192260.						
PR	24-JAN-1996;	96JP-0010365.						
PR	21-JUN-1996;	96JP-0161648.						
PR	24-JUN-1996;	96JP-0162813.						
PA	(TOYM) TOYO BOSEKI KK.							
PI	Shiba T, Taniguchi N, Uozumi N, Yanagidani S;							
DR	WPI: 1997-393690/36.							
DR	P-PSDB; AAW22125.							
XX	Human or pig alpha 1-6 fucosyltransferase and DNA encoding it - for							
XX	synthesis and modification of sugar chains and used as an antigen							
XX	for production of diagnostic antibodies							

P5	Claim 18, Page 39-43; 61pp; Japanese.
CC	AAT76573 and AAT76574 represent the coding sequences for the pig and
XX	human alpha 1-6 fucosyltransferases of the invention, respectively. The
CC	encoded enzyme transfers fucose from guanosine diphosphate to the
CC	6-hydroxyl group of the GlcNAc nearest to R in the receptor molecule:
CC	(GlcNAc)1-2Mannalpa 1-6)(GlcNAc)beta 1-2Mannalpa 1-3)(Mannobeta
CC	1-4)(GlcNAc)beta 1-4)(GlcNAc-R to give (GlcNAc)Mannabeta 1-2Mannalpa 1-6)(GlcNAc)beta
CC	1-Mannalpa 1-3)(Mannalpa 1-4)(GlcNAc)beta 1-4)(fucalpa 1-6)(GlcNAc-R. It has
CC	an optimum pH of about 7.0 (pig) or 7.5 (human), and is stable over the
CC	pH range 4-10 after 5 hours at 4 degrees C. The optimum working
CC	temperature of the enzyme is 30-37 degrees C. A bivalent metal is not
CC	required for activity of the enzyme, and the enzyme is not inhibited in
CC	the presence of 5 mM EDTA. The enzyme is useful in the synthesis and
CC	modification of sugar chains, and as antigen for the production of
CC	antibodies recognising the enzyme. The antibodies can be used for the
CC	diagnosis of cancer and other diseases.
XX	
SQ	Sequence 2100 BP; 651 A; 436 C; 489 G; 524 T; 0 other;
	Query Match 87.4%; Score 1510.4; DB 18; Length 2100;
	Best Local Similarity 92.1%; Pred. No. 0;
	Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
OY	1 ATGCGGCATGTAGCTGGTTCGTGGCCTTGAGATTATGCTCATTTCTTTTGGCTGGGGACC 60
Dd	198 ATGCGGCATGTAGCTGGTTCGTGGCCTTGAGATTATGCTCATTTCTTTTGGCTGGGGACC 257
OY	61 TTGCTATTTTACATAGGTGTCCTACTTGGTAGAGATTAAGCACTCTGATCACCTTAGC 120
Dd	258 TTGCTATTTTAAATAGGTGTCCTACTTGGTAGAGATTAAGCACTCTGATCACCTTAGC 317
OY	121 CGAGAACGTGTCACAATTTTGGCAAGCTGGAACCTTTAAAAACAATAATGAAGACTTG 180
Dd	318 CGAGAACGTGTCACAATTTTGGCAAGCTGGAACCTTTAAAAACAATAATGAAGACTTG 377
OY	181 AGGAGAATGGCTGATCTCTCCGAATPACACAGAAGGCCCATTTGATCAGGGGCCACCTTCA 240
Dd	378 AGGCGAATGGCGAATCTCTCCGATPACACAGAAGGCCCATTTGATCAGGGGCCACCTTCA 437
OY	241 GGAAGAGTTTCTGCTTTTGAAGAGCAATTTATGAAGGCCCAAAGACATTTGAAAATTTAT 300
Dd	438 GGAAGAGTTTCTGCTTTTGAAGAGCAATTTATGAAGGCCCAAAGACATTTGAAAATTTAT 497
OY	301 AAGAAACAACATTAANAATATGTCACAGGAAGGATCATGAATTCCTAAGAGAGAGATTGAA 360
Dd	498 AAGAAACAACATTAANAATATGTCACAGGAAGGATCATGAATTCCTAAGAGAGAGATTGAA 557
OY	361 AATGAGCTTAAGAAGACTCTGGTTTTTCTACAAAGTGAAGTGAAGAATTTAAGAATTTA 420
Dd	558 AATGAGCTTAAGAAGACTCTGGTTTTTCTACAGATGAATTTGAAGAATTTAAGAATTTA 617
OY	421 GAAGGAATTTGAATCTCAAGAAGATCAGATGAATTTCTATCAGATTTGGGACATCATGAA 480
Dd	618 GAAGGAATTTGAATCTCAAGAAGATCAGATGAATTTCTTTTGGATTTAGGACATCATGAA 677
OY	481 AGGCTTAAATAGAGGATCTATACACCTGATCAAGTCAAGAGATGGGGGAGGATTTGGCGT 540
Dd	678 AGGCTTAAATAGAGGATCTATACACCTGATCAAGTCAAGAGATGGGGGAGGATTTGGCGG 737
OY	541 GAAAGAGAGGCGCAAGATCTGCACAGAGCTGGTCCAGCGAGAAATTAACATATCTTCAGAA 600
Dd	738 GAAAGAGAGGCGCAAGATCTGCACAGAGCTGGTCCAGCGAGAAATTAACATATCTTCAGAA 797
OY	601 CCCAAGAGCTGCAGCAAAAGCCAAAGAGCTAGTGTGTAATATCAACAAAGCTGTGCGTAT 660
Dd	798 CCCAAGAGCTGCAGCAAAAGCCAAAGAGCTGTGTATATATCAACAAAGCTGTGCGTAT 857
OY	661 GGCTGTCAAGCTTCATCATATGATAGTACTAGTCTTTATGATTTGGCATATNGGACCCAGCAGACA 720
Dd	858 GGCTGTCAAGCTTCATCATATGATAGTACTAGTCTTTATGATTTGGCATATNGGACCCAGCAGACA 917
OY	721 CTCGGCTTGGAAATTCACAATTTGGCGCTACGGCTACTTGGGGGATGGGAACATGTGTTTGA 780

```
Db 918 CTCATCTTGGAAATTCAGAAATTCGGCTATGCTAGTGGATGGAGACTGATTTAGG 977
Qy 781 CCTGTAAGTGAACGTCGACAGACAGATCTGGCAGCTCCACTGGACATTTGCTAGTGA 840
Db 978 CCTGTAAGTGAACGTCGACAGACAGATCTGGCAGCTCCACTGGACATTTGCTAGTGA 1037
Qy 841 GTAAGGACAAAAATGTTAGAGGTGAGTCCCATTTGAGACAGTGTCTCTCTCT 900
Db 1038 GTGAAGGACAAAAATGTTAGAGGTGAGTCCCATTTGAGACAGTGTCTCTCTCTCT 1097
Qy 901 CCTCATATTTATCCCTGGCTGTCGACAGACCTTGCAGATCGACTGTAGAGTGCAT 960
Db 1098 CCTCATATTTATCCCTGGCTGTCGACAGACCTTGCAGATCGACTGTAGAGTGCAT 1157
Qy 961 GGTGATCTGTCAGTGTGGGTATCCCACTTTGTCAGTACTGATTTGCCACAAACC 1020
Db 1158 GGTGATCTGTCAGTGTGGGTATCCCACTTTGTCAGTACTGATTTGCCACAAACC 1217
Qy 1021 TGGCTGAAAAAGAAATAGAGAGGCGCACGAAGCTGAGGCTTCAACATCCAGTTATT 1080
Db 1218 TGGCTGAAAAAGAAATAGAGAGGCGCACGAAGCTGAGGCTTCAACATCCAGTTATT 1277
Qy 1081 GGAGTCATGTTAGAGCAGACAGAAAGTGGAGCGGAGCAGCCTTCCATCCATTTGAG 1140
Db 1278 GGAGTCATGTTAGAGCAGACAGAAAGTGGAGCGGAGCAGCCTTCCATCCATTTGAG 1337
Qy 1141 GAATACACGCTGACCTTGAAGACACTTTCACCTTCTGCTGCGAATGCAAGTGAAT 1200
Db 1338 GAATACACGCTGACCTTGAAGACACTTTCACCTTCTGCTGCGAATGCAAGTGAAT 1397
Qy 1201 AAAAAAAGGCTGATTTGGCCACAGATGACCCGCTTGTAAAGAGGCAAAAAAAG 1260
Db 1398 AAAAAAAGGCTGATTTGGCCACAGATGACCCGCTTGTAAAGAGGCAAAAAAAG 1457
Qy 1261 TACCCAGTATGAAATTTATGATGATTAATCTATCTTGTGACAGCTGACATTAAT 1320
Db 1458 TACCCAGTATGAAATTTATGATGATTAATCTATCTTGTGACAGCTGACATTAAT 1517
Qy 1321 CGATATACGAAATTTCACTTGGGGGTGATCTCGGATATACATCTTCTCCAGGCA 1380
Db 1518 CGATATACGAAATTTCACTTGGGGGTGATCTCGGATATACATCTTCTCCAGGCA 1577
Qy 1381 GACTTCCTGATGCTATTTTCATGTCAGAGTCTGTAAGGTTGCTATGAATCATGCA 1440
Db 1578 GACTTCCTGATGCTATTTTCATGTCAGAGTCTGTAAGGTTGCTATGAATCATGCA 1637
Qy 1441 GCGCTGATCTGATGCTCTGCAACTTCCGTTCTTTGGATGACATCTACTATTTTGA 1500
Db 1638 ACACATGATCTGATGCTCTGCAACTTCCGTTCTTTGGATGACATCTACTATTTTGA 1697
Qy 1501 GGGCCCAATGCCACACCAAAATGCAATTTATCTCCACCAACCTCGAAGTGAAGCA 1560
Db 1698 GGGCCCAATGCCACACCAAAATGCAATTTATGCTACCAACCTCGAAGTGAAGCA 1757
Qy 1561 ATCCCATGGAACCTGAGATATATTTGGTGGCTGGAATCACTGGATGGCTATTCCT 1620
Db 1758 ATCCCATGGAACCTGAGATATATTTGGTGGCTGGAATCACTGGATGGCTATTCCT 1817
Qy 1621 AAAGTGTTAACAGAAATCTGGAGAGAGCGGCTTATTCCTCTCAAAAGTTTCAGAG 1680
Db 1818 AAAGTGTTCACAGAAATTTGGAGAGAGCGGCTTATTCCTCTCAAAAGTTTCAGAG 1877
Qy 1681 AAGATAGAAACAGTCAAGTACCCACATATCCCGAGGCTGACAGATTA 1728
Db 1878 AAGATAGAAACAGTCAAGTACCCACATATCCCGAGGCTGACAGATTA 1925
```

```
RESULT 4
ABV22468
ID ABV22468 standard; cDNA: 3291 BP.
XX
AC ABV22468;
```

```
XX 13-SEP-2002 (first entry)
DT
XX Human prostate expression marker cDNA 22459.
DE
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene: ss.
XX Homo sapiens.
OS
XX WO200160860-A2.
PN
XX 23-AUG-2001.
PD
XX 20-FEB-2001: 2001WO-US05171.
PF
XX 17-FEB-2000: 2000US-183319P.
PR 16-MAR-2000: 2000US-189862P.
PR 25-MAY-2000: 2000US-207454P.
PR 09-JUN-2000: 2000US-211314P.
PR 18-JUL-2000: 2000US-219007P.
PR 13-DEC-2000: 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
XX WPI: 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
XX Claim 1: Page 3914; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 3291 BP; 925 A; 729 C; 745 G; 871 T; 21 other:
SQ
XX
XX Query Match 87.4%; Score 1510.4; DB 23; Length 3291;
XX Best Local Similarity 92.1%; Pred. No. 0;
XX Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
```

Qy	241	GGAAGACTGTGCTGTTTATGAAAGACAAATTTATGAAGGCCCAAGAACATTTGAAATTTAT	300
Dp	967	GGAAGACTGACCGCTTTTATGAGAGCAGCTTGTATAGCGCCAAAGAACAGTTGAAATTTAC	1026
Qy	301	AAGAAACAAACTAAAAATGCTCCAGGGAAGGATTCATGAATCTTAAGGAGAGAGATTGAA	360
Dp	1027	AAGAACAAGAACAGAAATGSGTCTGGGGGAAGGATTCATGAATCTTGAAGGAGGAGATTGAA	1086
Qy	361	AATGGAGCTAAAGACTCTGTGTTTTTTTCTACAAAGTGAAGTGAAGAAATTAAGAAATTTA	420
Dp	1087	AATGAGACTTAAGAGCTCTGTTTTTCTACAGAGTGAATTAAGAAATTTAAGAACTTTA	1146
Qy	421	GAAGGAATGACTCCCAAAACATGCAAGATGGAATTTCTCTCAGATTTGGGACATCATGAA	480
Dp	1147	GAAGGAATGAACTCCAAACATGCAAGATGAATTTCTTTTGATTTAGGACATCATGAA	1206
Qy	481	AGGCTATTAAATGACGAGTCTATCTACTCTCAGTCAACAGATGSGGGACGGTATGGCGT	540
Dp	1207	AGGCTATTAAATGAGGATCTATCTACTCTCAGTCAACAGATGAGAGAGAGTGTATGGCGG	1266
Qy	541	GAAGAAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCCGAGAAATTAATCTTCAGAA	600
Dp	1267	GAAGAAGAGGCCAAAGATCTGACAGAGCTGGTTCACCGGAGAAATTAATCTTCAGAA	1326
Qy	601	CCCAAGAGCTGACAGCAAGCCAAAGGCTAATGTTGTAATATCAACAAAGCTGTGGCTAT	660
Dp	1327	CCCAAGAGCTGACAGCAAGCCAAAGGCTGTAAATTAATCAACAAAGGCTGTGGCTAT	1386
Qy	661	GGCTGTAGGCTCCATCTATGTATGTATCTGCTGTTTATATATGATGTGGACCCAGGAACA	720
Dp	1387	GGCTGTAGGCTCCATCTATGTGTGTCTCTCTGCTTCATGATTTGCTATGGAACCCAGGAACA	1446
Qy	721	CTCGCCTTGGAAATCTCAAAATTTGGCCCTACGCTACTGGGGATGGGAAACTGTGTTTGA	780
Dp	1447	CTCATCTTGGAAATCTCAGATTTGGCCCTATGCTACTGGTGGATGGAGACTGTATTTAGG	1506
Qy	781	CCTGTAACTGAGAGCTGCACAGACACATCTGGCAGCTCCACTGTGACATTTGGTCAGGTAA	840
Dp	1507	CCTGTAACTGAGATATCACAGACACATCTGGCATCTCCACTGTGACACTGGTCAGGTAA	1566
Qy	841	GTTAAGGACAAAAATGTTTGAAGTGTGGAAGTCCCATTTGTAGACAGTTCATCTGCT	900
Dp	1567	GTTAAGGACAAAAATGTTCAAGTGTGTGAGCTTCCCATTTGTAGACAGTTCATCTGCT	1626
Qy	901	CTCTCATATTTTACCCCTGCTGTCCAGAAAGACTTGCAGATGCACTTGTACGAGTCCAT	960
Dp	1627	CTCTCATATTTTACCCCTGCTGTACAGAAAGACTTGCAGATGCACTTGTACGAGTCCAT	1686
Qy	961	GCTGATCTCTGCACTGTGTGTGGTATCCCAAGTTGTCAAGTACTTATGTTGCCACACACC	1020
Dp	1687	GCTGATCTCTGCACTGTGTGTGGTATCTCAAGTTGTCAAACTTATGATCCGCCACACAGCT	1746
Qy	1021	TGGCTGGAAGAAATAGAAAGAGGCCACCAAGAACTAGGCTTCAAAACATCCACTTAT	1080
Dp	1747	TGGCTGGAAGAAATAGAAAGAGGCCACCAAGAACTTGGCTTCAAAACATCCACTTAT	1806
Qy	1081	GGAGTCCATGTTTACAGCAGACAGAAAGTGGAGGCAAGCAAGCTTCCATCCCATTTGAG	1140
Dp	1807	GGAGTCCATGTCACAGCAGACAGAAAGTGGAAACGAAGTGCCTTCATCTCCATTTGAA	1866
Qy	1141	GAATACACGCTGACGCTTGAAGAGACTTTCAGCTTCTTGCTGCAGAAATGCAAGTGGAT	1200
Dp	1867	GAGTACATGCTGCTTGTGAAGAACTTTCAGCTTCTTGCAAGCAAGATGCAAGTGGAC	1926
Qy	1201	AAAAAAGGGGTATTGAGCAGAGATGACCCGTTTCTTAAAGAGGCAAAAAAAG	1260
Dp	1927	AAAAAAGAGGTATTGAGCAGAGATGACCCCTTCTTATTAAGAGGCAAAAAAAG	1986
Qy	1261	TACCCCAAGTTATGAATTTATTAAGTAACTATCTCTTGGTCAAGCTGCAGATACATAT	1320
Dp	1987	TACCCCAATTAATGAATTTATTAAGTAACTATCTATTTCTGCTCAAGCTGCAGATACAT	2046
Qy	1321	CGATATTCAGAAAAATTCATCTTGGGGGTGTGATCTGTGATATACACTTCTCTCCAGCA	1380

Db	2047	CGATACACAGAAATTTACCTTCCTGGAGAGTATCCTGGATATACATTTTCTCTCAGGCA	2106
Qy	1381	GACTTCTAGTGTGTACTTTTTCATCGCAGAGTCTGTAGAGTTGCTTATGAAATCATTGCAA	1440
Db	2107	GACTTCTAGTGTGTACTTTTTCATCGCAGAGTCTGTAGAGTTGCTTATGAAATCATTGCAA	2166
Qy	1441	GGCTGCATCCTGATGCTCTGGGAACCTCCGTTCTTTGGATGACATCTACTATTTTGGGA	1500
Db	2167	ACACTCATCTCTGTGTGCTCTGGAAACCTTCATCTTTTGTAGATGACATCTACTATTTTGGG	2226
Qy	1501	GGCCCAATGGCCCAACCAAAATTTGCATTTATCTCACCAACCTGAACTGAGAGAA	1560
Db	2227	GGCCAGAAATGCCCCACATCAATTTGCCATTTATGCTACACCAACCCGAACTGACATGAA	2286
Qy	1561	ATCCCCATGGAACCTGGAGATATTTATTTGGTGTGGCTGGAATTCACCTGGATGGCTATCT	1620
Db	2287	ATTCCTCATGGAACCTGGAGATATTTATTTGGTGTGGCTGGAATTCATTTGGATGGCTATCT	2346
Qy	1621	AAAGGTTTAAACAGAAACCTGGAGAGAGGCGCTTATTCCTCTACAAAGTTGCAAG	1680
Db	2347	AAAGGTTTCAACAGAAATTTGGAGAGGCGCTTATTCCTCTACAAAGTTGCAAG	2406
Qy	1681	AAAGTAAACAGTCAAGTACCCACATATCCGAGGCTGACAGTAA	1728
Db	2407	AAAGTAAACAGTCAAGTACCCACATATCTGAGGCTGAGAAATAA	2454
RESULT 5			
ABV28283			
ID	ABV28283	standard; cDNA: 3291 BP.	
XX	ABV28283:		
AC	16-SEP-2002	(first entry)	
DT			
XX	Human prostate expression marker cDNA 28274.		
DE			
XX	Human: prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;		
KW	pharmacogenomic marker; gene: ss.		
XX	Homo sapiens.		
OS			
XX	WO200160860-A2.		
FN			
XX	23-AUG-2001.		
PD			
XX	20-FEB-2001; 2001WO-US05171.		
PF			
XX	17-FEB-2000; 2000US-183319P.		
XX	16-MAR-2000; 2000US-189862P.		
PR	25-MAY-2000; 2000US-207454P.		
PR	09-JUN-2000; 2000US-211314P.		
PR	18-JUL-2000; 2000US-219007P.		
PR	13-DEC-2000; 2000US-255281P.		
XX			
XX	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.		
PA			
XX	Schlegel R, Endege WO, Monahan JE;		
PI			
DR	WPI: 2001-662795/76.		
XX			
PT	Novel isolated nucleic acid molecule associated with cancerous state of		
PT	prostate cells and correlating with presence of prostate cancer, useful		
PT	for detecting presence of prostate cancer, stage of prostate cancer -		
XX			
PS	Claim 1; Page 5883; 11750pp: English.		
CC	The invention relates to an isolated nucleic acid molecule (I) comprising		
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the		
CC	specification or its complement. (I) is useful for:		
CC	(a) assessing whether a patient is afflicted with prostate cancer;		
CC	(b) monitoring the progression of prostate cancer in a patient;		

xx	Homo sapiens.
os	WO200122920-A2.
xx	
xx	05-APR-2001.
xx	
xx	28-SEP-2000; 2000MO-US26524.
xx	
xx	29-SEP-1999; 99US-0157137.
xx	
xx	03-NOV-1999; 99US-0163280.
xx	
xx	(HUMA-) HUMAN GENOME SCI INC.
xx	
xx	
xx	Ruben SM, Barash SC, Birse CE, Rosen CA;
xx	
xx	WPI: 2001-235357/24.
xx	
xx	P-PSDB: AAG73884.
xx	
xx	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
xx	useful for preventing, diagnosing and/or treating colorectal cancers -
xx	
xx	Claim 1; Page 2480-2481; 9803pp; English.
xx	
xx	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
xx	cancer-associated nucleic acid molecules (N) and proteins (P), where
xx	the proteins are collectively known as colon cancer antigens. The colon
xx	cancer antigens have cytostatic activity and can be used in gene
xx	therapy and vaccine production. N and P may be used in the prevention,
xx	diagnosis and treatment of diseases associated with inappropriate P
xx	expression. For example, N and P may be used to treat disorders
xx	associated with decreased expression by rectifying mutations or deletions
xx	in a patient's genome that affect the activity of P by expressing of P.
xx	Inactive proteins or to supplement the patient's own production of P.
xx	Additionally, N may be used to produce the colon cancer-associated P,
xx	by inserting the nucleic acids into a host cell and culturing the cell
xx	to express the proteins. N and P can be used in the prevention, diagnosis
xx	and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
xx	and AAG77789 represent sequences used in the exemplification of the
xx	present invention.
xx	
xx	N.B. Pages 666 to 682 and page 7053 of the sequence listing were
xx	missing at time of publication, meaning no sequences are present for
xx	SEQ ID NO:1027 to 1052, 7921 and 7922.
xx	
xx	Sequence 3007 BP; 900 A; 597 C; 668 G; 834 T; 8 other;
xx	
xx	Query Match 86.7%; Score 1499; DB 22; Length 3007;
xx	Best Local Similarity 92.0%; Pred. No. 0;
xx	Matches 1591; Conservative 1; Mismatches 136; Indels 1; Gaps 1
0y	1 ATGCGCCATGAGACTGGTTCTGTCGTCCTGGATTATGCTCATTTCTTTGGCTGGGGACC 60
Db	420 ATGCGGCCATGAGACTGGTTCTGTCGTCCTGGATTATGCTCATTTCTTTGGCTGGGGACC 479
0y	61 TTGCGATTTTACATAGTGGTGCACCTTGGTAACGAGTAATGACACACCTGTATCACTAGC 120
Db	480 TTGCGATTTTATATAGTGGTGCACCTTGGTAACGAGTAATGACACATCTGTATCACTAGC 539
0y	121 CGAAGACTGTCACAGATTTTGGCAAGCTGGAACGCTTTAAACACAAATTAAGAACTTG 180
Db	540 CGAAGACTGTCACAGATTTTGGCAAGCTGGAACGCTTTAAACACGAGATTAAGAACTTG 599
0y	181 AGGAGATGCGTGG-CATCTCTCCGAATACCAAGAGGCCCATTTGATCAGGGCCAGCTTC 239
Db	600 AGGCAATGCGCGAAATCTCTCCGATACAGAGGCCCATTTGATCAGGGCCAGCTTAT 659
0y	240 AGGAGAGTGTGCGCTTTGAAGAGCAATTTTAAAGAGGCCCAAGAACAGATTTGAAATTA 299
Db	660 AGGAGAGTGTGCGCTTTGAAGAGCAATTTTAAAGAGGCCCAAGAACAGATTTGAAATTA 719
0y	300 TAAGAAACAACTAAATAATGTCTCAGGAGAGCATCATGAATCTTAAGGAGGAGATTGA 359
Db	720 CAAAGAAACAGACAAATAATGTCTGAGGAGAGCATCATGAATCTTGAAGGAGGAGATTGA 779

QY	360	AAATGAGACTAAAGAGCTCTGCTGTTTTCTTCTACAAAGTGAAGTGGAGAATTTAAACAATTTT	413
Db	780	AAATGAGACTAAAGAGCTCTGCTGTTTTCTTCTACAGAGATGAAATGGAAACAAATTTAAACAATTTT	839
QY	420	AGAAGGAATGAACTCCCAAGACATGCAGATGGAATTTCTATCAAGATTTTGGACATCATGA	479
Db	840	AGAAGGAATGAACTCCCAAGACATGCAGATGGAATTTCTTTTGGATTTAGGACATCATGA	899
QY	480	AAGCTTATTAATGACGATCTATTAACCTCACTCAACAGATGGGCGAGTGTATTTGGCG	539
Db	900	AAGGCTTATTAATGACGAGTCTATTAACCTCACTCAACAGATGAGGAGCGAGTGTATTTGGCG	959
QY	540	TGAAAAGGAGGCGCAAAACATCTGCAGAGCGGTGGTCCAGCGGAGAAATTAACATATCTTTCAGAA	599
Db	960	GGA AAAAGAGGCGCAAAACATCTGCAGAGAGCTGGTTCACGCGGAGAAATTAACATATCTTTCAGAA	1019
QY	600	TCCCAAGAGCTGCAGCAAAAGCCAAAGAGCTAGTGTATTAATTAACAACAAAGCGCTGTGGCTA	659
Db	1020	TCCCAAGAGCTGCAGCAAAAGCCAAAGAGCTAGTGTATTAATTAACAACAAAGCGCTGTGGCTA	1079
QY	660	TGGCTGTACGCTCCATCATATGATGTAAGTCTTTATGATTTGCATATGGCACCCACCGAAC	719
Db	1080	TGGCTGTACGCTCCATCATATGATGTAAGTCTTTATGATTTGCATATGGCACCCACCGAAC	1139
QY	720	ACTGCGCTTGGAAATCTCAAAATTTGGGCGCTACGCTACCTGGGGGATGGGAAATCTGTATTAG	779
Db	1140	ACTCATCTTGGAAATCTCAAAATTTGGGCGCTATGCTACTGGTGATGGGAGACTGTATTTAG	1199
QY	780	ACCTGTAAAGTAGAGCGTGCACAGACAGATCTGGCAGCTCCACTGGAATTTGGTCAGGTGA	839
Db	1200	GCCCTGTAAAGTAGAGCATGCACAGACAGATCTGGCATCTCCACTGGCACCTGGTCAGGTGA	1259
QY	840	AGTAAAGGACAAAATGTTCAAGGGGTGGAGTGGCGCCCATTTGTAAGACAGTTCATCACTCCG	899
Db	1260	AGTGAAGGACAAAATGTTCAAGGGGTGGAGTGGCGCCCATTTGTAAGACAGTTCATCACTCCG	1319
QY	900	TCCTCCATATTTACCCTCGCTGTCCAGAAAGACCTTGCAGATTCGATCTGTACGAGTCCA	959
Db	1320	TCCTCCATATTTACCCTTGGCTGTACAGAAAGACCTTCGAGATTCGATCTGTACGAGTCCA	1379
QY	960	TGGTGTATCTCTGACAGTGGTGGGTATCCCAAGTTTGTCAAGTACTTGTATTCGGCCCAACAC	1019
Db	1380	TGGTGTATCTCTGACAGTGGTGGGTATCTCAAGTTTGTCAAAATCTGTATTCGGCCCAACAC	1439
QY	1020	CTGCGCTGGAAAAGGAATTAAGAAGAGCCACAAGAGCTAGAGCTTCAAAATCCACAGTTAT	1079
Db	1440	TTTGCTGTAAAAGAAATTAAGAAGAGCCACAAGAGCTTGGCTTCAAAATCCACAGTTAT	1499
QY	1080	TGGAGTCCATTTAGACGCACAGACAAAGTGGGAGCGGAAGCCCTTCATCCCATTTGA	1139
Db	1500	TGGAGTCCATTTAGACGCACAGACAAAGTGGGAGCAGAAAGCTGCTTCATCCCATTTGA	1159
QY	1140	GGAATFACACGGTGCAGCTTGAAGAAGCTTTCAGCTTCTTGTCGCGAGAAATTCAGTGTGA	1199
Db	1560	AGAGTACATGCTGCATTTGAAGAACATTTTCAGCTTCTTGCAACGCGAGATTCAGTGTGA	1619
QY	1200	TAAAAAAAAGGGTGTATTTGGCCACAGATGACCCCTGCTTTGTTAAAAAGAGCAAAAACAAA	1259
Db	1620	CAAAAAAAGAGTGTATTTGGCCACAGATGACCCCTTCTTTATTAAGAAGAGCAAAAACAAA	1679
QY	1260	GTACCCAGTTATGAATTTATTAAGTATACTTATCTCTTGGTCAAGCTGACTACATTA	1319
Db	1680	GTACCCAGTTATGAATTTATTAAGTATACTTATTTCTGGTCAAGCTGACTGACAA	1739
QY	1320	TCGATATTAACAGAAAATTCACCTTGGGGGTGATCCCGAATTAACACTTCTCTCCCAAGC	1379
Db	1740	TCGATACACAGAAAATTCACCTTGGGAGTGAATCCCTGGATATACATTTTCTCTCAGGC	1799
QY	1380	AGACTTCCATGTGTACTTTTTCATCGCAGGCTGTGTAGAGTTGCTTATGAATATCATCA	1439
Db	1800	AGACTTCCATGTGTACTTTTTCATCCAGGCTGTGTAGAGTTGCTTATGAATATCATCA	1859

Dh	1060	GGATATCTCGCAGTGTGGTGGGTATCCCAAGTTTCTCAATATCTTGATCCGTCCCAACT	1119
Qy	1021	TGCGTGGAAAAGCAATATGAAAGAGCCCAACAAGCTTAGGCTTCAAACTCCAGTTATT	1080
Dh	1120	TGGCTGGAAAAGGAATATGAGAAACCAACCAAGAGCTTGGCTTCAAACTCCAGTTATT	1179
Qy	1081	GGAGTCCATGTTTAGACGCACAGCAAAAGTGGGAGCGGAGACGCTTCCATCCATTGAG	1140
Dh	1180	GGAGTCCATGTCAGACGCACATGCAAAAGTGGGAACAAGAGGCTTCCATCCATTGAG	1239
Qy	1141	GAATACACGGTGGTGAAGAAGACTTTAGCTTTGCTGCTGCAGATATGCAAGTGAT	1200
Dh	1240	GAATACATGATGATACACGTTGAAGAACTTTTACGCTTCTCGAACCAATATGAATGGAT	1299
Dh	1300	AAAAAAGATGATATCTGGCCACATGATAGACCTTCTTGTGTTAAAGAGGCAAGACAAAG	1359
Qy	1261	TACCCCACTTATGAATTTTATTAAGTATACCTATATCTTTGGTCAGCTGGACTACATAT	1320
Dh	1360	TACTCCAAATTAATGAATTTATTAAGATTAACATCTTTTCTTGGTCAGCTGGACTACAAAC	1419
Qy	1321	CGATTAACAGAAAATATCACTTGGGGGTGTATCCGTGATATACACTTCTCTCCACAGCA	1380
Dh	1420	CGATTAACAGAAAATATCACTTGGGGGTGTATCCGTGATATACACTTCTCTCCACAGCT	1479
Qy	1381	GACTTCCTAGTGTGACTTTTTCATCGCAGAGCTGTAGAGTGGCTATGAATCATGCAA	1440
Dh	1480	GACTTCCTGTGTGTACTTTTTCATCCAGGTGTGTAGGGTGGCTATGAATCATGCAA	1539
Qy	1441	GGCGTGCATCTGATGCTCTTGCGAACTTCCGTTCTTTGGATGACATCTACTATTTTGA	1500
Dh	1540	ACACGTGCATCTGTATGCTCTGCAAACTCCATCTTATGATGACATCTACTATTTTGA	1599
Qy	1501	GGCCCAATATGCCAACCAAAATGGCAATTAATCTCAACAACAACCTCGAAGCAAGAGAA	1560
Dh	1600	GGCCAAATATGCCAACCAACAATGCACTTATATCTCAACAACCTCGAAGCAAGAGAGAA	1659
Qy	1561	ATCCCATGGAACCTGGAGATATTTATTTGGTGTGGCTGGAAATCACTGGATGGCTATCT	1620
Dh	1660	ATCCCATGGAACCTGGAGATATTCATTTGGTGTGGCTGGAAACCATTTGGAATGGTACTCT	1719
Qy	1621	AAAGTGTTAAACAGAAAATCTGGGAAGGACGGGCTATATCTCTCTCAAAATGTTGAGAG	1680
Dh	1720	AAAGTGTTCAACAGAAAATCTAGGAAAAACAGGCTGTATACCTTCTCAAAAGTCCGAGAG	1779
Qy	1681	AAAGATGAAGACGTCAAGTACCCCAATATCCGAGGCTGCAAGTA	1727
Dh	1780	AAAGATGAAGACGTCAAGTACCCCAATATCTCAAGCTGAAAAATA	1826
RESULT 8			
ABK70032			
ID	ABK70032 standard; cdNA: 1728 BP.		
XX	ABK70032:		
AC	15-JUL-2002 (first entry)		
XX	Antibody production method related cdNA #2.		
DE	Antibody production: cytosstatic; immunomodulator; vasotropic; virucide;		
XX	antibacterial; antiinflammatory; antiallergic; allergy; inflammation;		
KW	autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;		
KX	tumour; circulatory disease; Infection; primer: ss.		
XX	Mus musculus.		
OS	WO200231140-A1.		
PN	18-APR-2002.		
PD	05-OCT-2001; 2001WO-JP08804.		
XX			
PE			

XX 06-OCT-2000; 2000JP-0308526.
PR
XX
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
XX
PI Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;
PI Hosaka E, Yamano K, Yamasaki M, Hanai N;
XX
XX
DR WPI: 2002-340182/37.
XX
XX
PR Cells producing antibody compositions including antibody fragments and
PR fusion proteins with Fc domain of antibody, useful for prevention or
PR treatment of cancer, immune diseases, circulatory diseases and
PR infections -
XX
XX
PS Claim 32: Page 4-11; 314pp; Japanese.

CC This invention relates to novel method for antibody production
CC comprising a Chinese hamster ovarian tissue-originated (CHO) cell
CC transferred with a gene encoding an antibody molecule for producing a
CC composition comprising an antibody molecule with an Fc domain bonded
CC to the N-glycoside linkage complex sugar chain. The produced antibody
CC compositions are drugs for prevention or treatment of diseases
CC accompanying tumour, allergy or inflammation, autoimmune diseases,
CC circulatory diseases, and viral and bacterial infections. The
CC antibodies can be stably produced using the method of the invention
CC with high binding activity and potency thus leading to high safety and
CC reduced side effects when applied alone or in combination with other
CC drugs for therapy. The present sequence represents a nucleotide
CC molecule used in the method of the invention.

Sequence 1728 BP; 529 A; 365 C; 400 G; 434 T; 0 other;

Query Match	84.98;	Score 1467.8;	DB 24;	Length 1728;
-------------	--------	---------------	--------	--------------

Matches 1565; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Qy	1	ATGGCGGCATGAGCTGGTCTGGCGGCTGGATTTATGCTCATCTTTTTCGCTGGGGACC	60
Db	1	ATGCGGGATGGAGCTGGTCTGGCGGCTGGATTTATGCTCATCTTTTTCGCTGGGGACC	60
Qy	61	TTGCTATTTTACATAGAGTGGTCACTTGGTACGAGATATAGCCACTCTGATCACTAGC	120
Db	61	TTGCTATTTTATATAGAGTGGTCACTTGGTATGAGATATAGCCACTCTGATCACTAGC	120
Qy	121	CGAGACCTGTCCAGATTTTGGCCAAGCTGGAACGCTTAAAAACACAATTTGAAGACTTG	180
Db	121	AGAGAACCTCTCAAGATTTCTTGGCAAAAGCTTGAAAGCTTTAAAAACGCAAAATGAAGACTTG	180
Qy	181	AGGAAATAGCGTGGATCTCTCCGATACCAAGAGCCCCATTTGATCAGGGGCCAGCTTCA	240
Db	181	AGGCAATAGCGTGGATCTCTCCGAATACCAAGAGCCCCATTTGATCAGGGGACAGCTACA	240
Qy	241	GGAGAGTTCTGCTTTTGAAGAGCAATTTATGAAGCCAAAGAACAGATTGAAATTTAT	300
Db	241	GGAAAGTCCCTGTTTATGAAGAACAGCTTTGTAAAGCCAAAGAACAGATTGAAATTTAC	300
Qy	301	AAGAAACAACCTAAATAATGTGTCCAGGGAAGGATCATGAATTCCTAAAGSAGAGATTGAA	360
Db	301	AAGAAACAACCTAAGAAATGTGTCTGGGAAGGATCATGAATTCCTAAAGAGAGATTGAA	360
Qy	361	AATGAGCTAAAGAGCTGTGTTTTTCTACAAAGTAGATTGAAGAAATTTAAAGATTTTA	420
Db	361	AATGAGCTAAAGAGCTGTGTTTTTCTACAAAGCGCAACGAAAGAAATTTAAAGCATTTTA	420
Qy	421	GAAGGAATTTGAATCCAAAGACATGACAGATCAATTTCTATCAGATTTTGGGACATCATGAA	480
Db	421	GAAGGAATTTGAATCTCAAGACATGACAGATCAATTTCTTTTGGATTTTAGCACACATGAA	480
Qy	481	AGCTTATATATGAGGATCTATACTACCTCAGTCAAAACAGATGGGGCAGGTGATTGGCGT	540
Db	481	AGGCTATATATACAGATCTATACTACTCAGTCAAAACAGATGGAGCAGGGGATTTGGCGT	540

```
QY 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCACAGCGAGAAATATCATATCTTCAGAAAT 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 541 GAAAAAGAGCCAAAGATCTGACAGAGCTGTCACAGCGAGAAATATCATATCTTCAGAAAT 600
QY 601 CCCAAGAGCTGACGACAAAGCCAAAGAGCTAGTGTATATATCAACAAGAGCTGTGCTAT 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 601 CCTAAGAGCTGACGACAAAGCCAAAGAGCTAGTGTATATATCAACAAGAGCTGTGCTAT 660
QY 661 GGCTGCTACCTCCATCATATAGTGTACTGCTTATATATGCAATATGGCACCCAGCCAGACA 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 661 GGTTGTCAACTCCATCATCTGCTACTGCTTATATATGCTTATGCAACCCAGCCAGACA 720
QY 721 CTCGCTTGGAAATCTCACAAATTTGGCGCTACGCTACGCGGAGTGGAAACTGTTTAGA 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 721 CTCATCTTGGAAATCTCACAAATTTGGCGCTACGCTACGCTGAGTGGAGAGCTGTTTAGA 780
QY 781 CCTGTAAGTAGAGAGCTGACAGACAGATCTGGCAGCTCCAGCTGAGATTGGTCAAGSTGAA 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 781 CCTGTAAGTAGAGAGCTGACAGACAGATCTGGCAGCTCCAGCTGAGATTGGTCAAGSTGAA 840
QY 841 GTAAAGGACAAAAAATTTTCAGGTGGTTGAGCTCCCATTTGTAGACAGTGTCTCCTCGT 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 841 GTAAATGACAAAAAATTTTCAGGTGGTTGAGCTCCCATTTGTAGACAGCTCCTCCTCGG 900
QY 901 CCTCCATTTTACCCCTGGCTGTCCAGAAAGACCTTGACAGATGCACTTGTAGAGATCCAT 960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 901 CCTCCTTACTTACCACTGGCTGTTCAGAAAGACCTTGACAGATGCACTTGTAGAGATCCAT 960
QY 961 GGTGATCTGACAGTGTGGTGGTATCCAGTTTGTCAAGTACTTGTATTCGCCCAACCC 1020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 961 GGTGACCTCTGACAGTGTGGTGGTATCCAGTTTGTCAAGTACTTGTATTCGCCCAACCC 1020
QY 1021 TGGCTGGAAAAAGAAATTAAGAGAGCCACCAAGAGCTTACGCTTCAACATCCAGTTAT 1080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1021 TGGCTGGAAAAAGAAATTAAGAGAGCCACCAAGAGCTTACGCTTCAACATCCAGTTAT 1080
QY 1081 GGAGTCCATGTAGAGCAGACAGAAAGTGGGAGCGGAAGCAGCCCTCCATCCCATTTGAG 1140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1081 GGAGTCCATGTAGAGCAGACAGAAAGTGGGAGCGGAAGCAGCCCTCCATCCCATTTGAG 1140
QY 1141 GAATACAGCGTGACAGCTTGAAGAAGACTTTCAGCTTCTTGCTCGAGAAATGCAAGTGAT 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1141 GAATACAGCGTGACAGCTTGAAGAAGACTTTCAGCTTCTTGCTCGAGAAATGCAAGTGAT 1200
QY 1201 AAAAAAGGCTGATTTGGCCACAGATGACCCCTGCTTTGTTAAAGAGCGAAAAAACAAG 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1201 AAAAAAGGCTGATTTGGCCACAGATGACCCCTGCTTTGTTAAAGAGCGAAAAAACAAG 1260
QY 1261 TACCCAGTTATGAATTTTATAGTGAATCTATCTTGTGACAGCTGACATACATTAAT 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1261 TACCCAGTTATGAATTTTATAGTGAATCTATCTTGTGACAGCTGACATACATTAAT 1320
QY 1321 CGATATACAGAAAAATTCACCTGGGGGTGTGATCCTGATATACACTTTCTCCACAGCA 1380
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1321 CGATATACAGAAAAATTCACCTGGGGGTGTGATCCTGATATACACTTTCTCCACAGCA 1380
QY 1381 GACTTCTAGTGTGACTTTTTCATGCGAGGTCTGTAGAGTTGCTTATGAATCATGCAA 1440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1381 GACTTCTAGTGTGACTTTTTCATGCGAGGTCTGTAGAGTTGCTTATGAATCATGCAA 1440
QY 1441 GGCGTGCATCTGATGCTCTGCGAAGCTTCCGTTTGGATGACATCTACTATTTTGA 1500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1441 GGCGTGCATCTGATGCTCTGCGAAGCTTCCGTTTGGATGACATCTACTATTTTGA 1500
QY 1501 GGCCCAAAATGCCCAACAATTTGCCATTTATCTCACCACCTCGAAGTGAAGAGAA 1560
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1501 GGCCCAAAATGCCCAACAATTTGCCATTTATCTCACCACCTCGAAGTGAAGAGAA 1560
QY 1561 ATCCCATGGAACCTGGAATATATTGTTGGTGGCTGGAATACATCGGATGCTATCCT 1620
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1561 ATTCATGGAACCTGGAATATATTGTTGGTGGCTGGAATACATCGGATGCTATCCT 1620
QY 1621 AAGGTGTTAAACAGAAAGCTGGGAAGAGCGGCTATATCCCTCTTCAAAAGTTGAGAG 1680
```

```
DB 1621 AAGGTGTTAAACAGAAAGCTGGAAAAACAGCGCTTATATCTCCCTCAGAAAGTCCAGAG 1680
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1681 AAGATGAAAGAGTCAAGTACCCACATATTCGCGAGCTGACAGATA 1727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1681 AAGATGAAAGAGTCAAGTATCCACATATCTGGAAGCTGAAAAATA 1727
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
AA87952
ID AAF87952 standard; DNA; 1017 BP.
XX
AC AAF87952;
XX
DT 20-JUL-2001 (first entry)
XX
DE Human alpha 1-6 fucosyltransferase 237-575 encoding DNA SEQ ID NO:1.
XX
KW Human: alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
KW alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunosay;
KW ds.
XX
OS Homo sapiens.
XX
FH Key 1.1017 location/Qualifiers
FT CDS /partial a
FT /product= "alpha 1-6 fucosyltransferase 237-575"
PN JP2001011097-A.
XX
PD 16-JAN-2001.
XX
PF 29-JUN-1999: 99JP-0183569.
XX
PR 29-JUN-1999: 99JP-0183569.
XX
PA (FJRE ) FUJIREBIO KK.
XX
DR WPI: 2001-275926/29.
XX P-PSDB: AAB75061.
XX
PT Novel anti-human alpha1-6 fucose transferase antibody useful for
PT immunosassay -
PS Example 1: Page 6-7; 11pp; Japanese.
XX
CC The present invention describes an anti-human alpha 1-6
CC fucosyltransferase (alpha 1-6 fucT, also called alpha 1-6 fucose
CC transferase) antibody. Also described are: (1) a hybridoma producing
CC the above monoclonal antibody; (2) an immunosassay for detecting human
CC alpha 1-6 fucT by using the above antibody of its antibody fragment;
CC and (3) a reagent used for the above immunosassay. The anti-human
CC alpha 1-6 fucT antibody can be used for immunosassay. The present
CC sequence encodes the human alpha 1-6 fucT protein of residues 237 to
CC 575, which is used in an example from the present invention.
XX
SQ Sequence 1017 BP; 296 A; 231 C; 228 G; 262 T; 0 other;

Query Match 50.8%; Score 878.4; DB 22; Length 1017;
Best Local Similarity 91.5%; Pred. No. 5.6e-241;
Matches 930; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 709 ACCCAGCAACTCGCTTGAATCTCAATTTGGCGCTACGCTAGTGGGAGTGGAA 768
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 ACCCAGCAACTCGCTTGAATCTCAAGAAATTTGGCGCTACTGCTGATGGAG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 769 ACTGTGTTAGACCTTAAAGTGAAGAGCTGACAGACAGATCTGGCAGCTCCACAT 828
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 ACTGTATTTAGGCTCTAAGTGAAGAGCTGACAGACAGATCTGGCAGCTCCACAT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 829 TGGTCAAGTGAAGTAAAGCAAAAATGTTCAAGTGTGAGCTCCCATTTGAGACAGT 888
```

```
Db 121 TGGTCAGGTGAAGTGAAGACAAATGTTCAGAGTGTGAGCTTCCATTGTAGACAGT 180
Oy 889 GTTTCATCTGTCCTCCCATATTTACCCGTGGCTGCCAGAAACCTTGACATTCACATT 948
Db 181 CTTTCATCCCGTCTCTCATATTTTACCTTGGCTGTACAGAAACCTCGACATTCACATT 240
Oy 949 GTACGAGTCATGTCATCTGTCAGTGTGGGTATCCAGTTTGTCAAGTACTTGATT 1008
Db 241 GTACGAGTCATGTCATCTGTCAGTGTGGGTATCCAGTTTGTCAAGTACTTGATT 300
Oy 1009 CGCCACAAACCTGTGCGAAGAAATAGAAAGAGCCACCAAGAGCTTGGCTTCAA 1068
Db 301 CGCCACAAACCTGTGCGAAGAAATAGAAAGAGCCACCAAGAGCTTGGCTTCAA 360
Oy 1069 CATCCAGTTATTTGAGATTCATGTTAGACGACACCAAAAGTGGAGGAGAGGCTTTC 1128
Db 361 CATCCAGTTATTTGAGATTCATGTTAGACGACACCAAAAGTGGAGGAGAGGCTTTC 420
Oy 1129 CATCCAGTTGAGAAATACACGTCACGTCGTCGAGAACACTTTCAGCTTCTGTCGACAG 1188
Db 421 CATCCAGTTGAGAAATACACGTCGTCGAGAACACTTTCAGCTTCTGTCGACAG 480
Oy 1189 ATGCAAGTGAATTAATAAAGGCTGTATTTGGCCACAGATGACCCCTTCTTATTAAGAG 1248
Db 481 ATGCAAGTGAATTAATAAAGGCTGTATTTGGCCACAGATGACCCCTTCTTATTAAGAG 540
Oy 1249 GCATAAACAAGTACCCAGTTATGAAATTTATTTAGTATACCTTCTGTCGTCGAGCT 1308
Db 541 GCATAAACAAGTACCCAGTTATGAAATTTATTTAGTATACCTTCTGTCGTCGAGCT 600
Oy 1309 GGAATCATATATGATATACAGAAATTCACCTTGGGGTGTGATCCTGATATACACTTT 1368
Db 601 GGAATCATATGATATACAGAAATTCACCTTGGGGTGTGATCCTGATATACACTTT 660
Oy 1369 CTCTCCAGGACAGCTTCTGATGTGTACTTTTCAATCCAGGTCGTAGAGTTGCTTAT 1428
Db 661 CTCTCCAGGACAGCTTCTGATGTGTACTTTTCAATCCAGGTCGTAGAGTTGCTTAT 720
Oy 1429 GAAATCATGCAAGGCTGATCTGATGCTGCGAAGCTTCCGTTCTTGGATGACATC 1488
Db 721 GAAATCATGCAAGGCTGATCTGATGCTGCGAAGCTTCCGTTCTTGGATGACATC 780
Oy 1489 TACTATTTTGAAGCCCAATGCCCACAAACCAATTTGCCATTTCTGACCAACCTCGA 1548
Db 781 TACTATTTTGAAGCCCAATGCCCACAAACCAATTTGCCATTTCTGACCAACCTCGA 840
Oy 1549 ACTGAAGAGAAATTCCTCATGGAACCTGGAGATATTTATGCTGTGCTGGAATTCAC 1608
Db 841 ACTGAAGAGAAATTCCTCATGGAACCTGGAGATATTTATGCTGTGCTGGAATTCAC 900
Oy 1609 GATGGCTATCTTAAAGGTGTACAGAAACCTGGAAAGGCGGCGCTATATCCCTCTAC 1668
Db 901 GATGGCTATCTTAAAGGTGTACAGAAACCTGGAAAGGCGGCGCTATATCCCTCTAC 960
Oy 1669 AAAGTTTCGAGAGAGATAGAAACAGTCAAGTACCCACATATCCGAGGCTGACAA 1724
Db 961 AAAGTTTCGAGAGAGATAGAAACAGTCAAGTACCCACATATCCGAGGCTGACAA 1016

RESULT 10
AAC63891
ID AAC63891 standard: cdna: 979 bp.
XX
AC AAC63891:
XX
DT 09-FEB-2001 (first entry)
XX
DE Chinese hamster FUT8 cDNA.
XX
KW Immunologically functional molecule; immune system; immunomodulation;
glycosylation; fucose; N-acetylglucosamine; cancer; circulatory disease;
viral infection; bacterial infection; allergy; autoimmune disease;
KW
```

```
KW Inflammation; antibody; Chinese hamster; Fut8; ss.
XX
OS Cricetus griseus.
PN WO200061739-A1.
PD 19-OCT-2000.
PE 07-APR-2000; 2000MO-JP02260.
PR 09-APR-1999; 99JP-0103158.
XX (KYOW ) KYOMA HAKKO KOGYO KK.
PI Hanai N, Nakamura K, Shoji E, Yamasaki M, Uchida K, Shinkawa T;
PI Imabeppu S, Kanda Y, Yamane N, Anazawa H;
DR WPI: 2000-665129/64.
XX
PT Control of activity of antibodies and other immunologically functional
PT molecules by addition or removal of fucose from sugar chain for
PT diagnosis and treatment of cancer, allergy and other diseases
XX
PS Example 8; Page 72-73; 81pp; Japanese.
XX
CC The invention relates to a method for controlling the activity of an
CC immunologically functional molecule (e.g., an antibody) where the
CC control is effected by the presence or absence of fucose bound to an
CC N-acetylglucosamine residue at the reducing end of the sugar chain on
CC the immunologically functional molecule. The invention also relates to
CC methods for the diagnosis, prevention or treatment of diseases which
CC involve the modified immunologically functional molecule, and agents
CC which stimulate the activity of an immunologically functional molecule.
CC The methods of the invention are used for the diagnosis, treatment and
CC prevention of a broad range of diseases including cancer, circulatory
CC disease, viral or bacterial infection, allergy, autoimmune disease and
CC inflammation. The present sequence represents a Chinese hamster Fut8
CC cDNA isolated in an exemplification of the invention.
XX
SO Sequence 979 bp; 286 A; 227 C; 218 G; 248 T; 0 other;

Query Match 48.2%; Score 833.4; DB 21; Length 979;
Best Local Similarity 90.7%; Pred. No. 4,3e-228;
Matches 888; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

Oy 720 ACTGCGCTTGGAATTCACAAATGGCGCTACGCTACCTGGGGGATGGGAAACGTGTAG 779
Db 1 ACTCATCTTGGAAATCTCAGAAATGGCGCTACGCTACGAGATGGAGACTGTGTAG 60
Oy 780 ACCTGTAAGTGAAGCTGACACAGACAGATCTGACACCTCCACTGACATTTGTCAGGTGA 839
Db 61 ACTGTAAGTGAAGCAATGACACAGACAGAGTCTGCGCTCCACATGAGACATGTCAGGTGA 120
Oy 840 ACTGAAGCAAAATGTTCAAGTGTGAGTGTGACCTCCATTTGTAGACAGTGTATCTCTG 899
Db 121 AGTGAAGGACAAATATGTTCAAGTGTGAGTGTGACCTCCATTTGTAGACAGTGTCTCTG 180
Oy 900 TCCCTCATATTTTACCCCTGCGCTGCCAGAAAGCTTGGAGATGCACTTGTACAGGTCA 959
Db 181 TCCCTCATATTTTACCCCTGCGCTGCCAGAAAGCTTGGAGATGCACTTGTACAGGTCA 240
Oy 960 TGGTGAATCTGCAAGTGTGCTGATCCAGTTTGTCAAGTACTTGATTCGCCCAAC 1019
Db 241 TGGTGAATCTGCAAGTGTGCTGATCCAGTTTGTCAAGTACTTGATTCGCCCAAC 300
Oy 1020 CTGCTGGAAGAAATAGAAAGAGGCGCACCAAGAGCTTGAAGCTTCAAAATCAGTTAT 1079
Db 301 TTGGCTGGAAGGAAATAGAAAGAAACCAAGAGAGCTTGGCTTCAAAATCAGTTAT 360
Oy 1080 TGGAGTCCATGTTAGACGACACAGCAAAAGTGGAGGGAAGCAAGCTTCATCCATTTGA 1139
Db 361 TGGAGTCCATGTTAGACGACACAGCAAAAGTGGAGGGAAGCAAGCTTCATCCATTTGA 420
```


Db	661	TGACTTCCTGTGTGTACTTTTTCATCCAGAGCTGTAGGTTTGCTATGAAATCATGTCA	720
Oy	1440	AGCCCTGCATCTGTGATGCCCTCTCGCAACTTCCCTTCTTGATGATGACATCTACTATTTGG	1499
Db	721	AACATCTGCATCTGTGATGCCCTCTCGCAAACTTCCATCTTTAGATGACATCTACTATTTTGG	780
Oy	1500	AGGCCCAATATCCGACACCAACCAATTTGTCATTTATCTCACCAACCTCGAATCGAAGSAGA	1559
Db	781	AGGCCCAATATCCGACACCAACCAAGATTTGCAAGTTTATCTCTCACCAACCTCGAATCAAGSAGA	840
Oy	1560	AATCCCATGGAACCTTGAGATATTTATGTGTGTGCTGGAATTCACCTGGATGCTATTC	1619
Db	841	AATCCCATGGAACCTTGAGATATTCATTTGTGTGTGCTGGAACCAATTTGGAATGCTTACTC	900
Oy	1620	TAAAGTGCTTAAAGAAAGCTGGGAAAGAGGCGCTATATCCCTCTACAAAGTTTCGAGA	1679
Db	901	TAAAGTGCTTAAAGAAAGCTTAGSAAAMAAAGCGCTGTACCTTCTTCAAAAGTCCGAGA	960
Oy	1680	GAGATAGAAAACAGTCAAG	1698
Db	961	GAGATAGAAAACGCTCAAG	979

CC	in an exemplification of the invention.
XX	
5Q	Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other

Query Match	47.98;	Score 827;	DB 21;	Length 979;
Best Local Similarity	90.38;	Pred. No. 2.9e-226;		
Matches 884;	Conservative 0;	Mismatches 95;	Indels 0;	Gaps 0;

QY	720	ACGGCCCTTGGAAATCTCCAAAAATTTGGGGCTTAACGCTACTGGGGGATGGGAACCTGTCTTTTAA	779
Db	1	ACTCATCTTTGGAAATCTAGAAATTTGGGGCTTAAGTACTGGTGGATGGGAAGACTGTCTTTAG	60
QY	780	ACCTGTAAGTGAGACGTGCACAGACAGATCTGGACAGCTCCACTGGACAATTTGGTCAGGTGA	839
Db	61	ACCTGTAAGTGAGACATGCACAGACAGATCTGGCCCTCTCCACTGGGACACTGGTCAGGTGA	120
QY	840	AGTAAAGCACAATAATGTTCAAGTGGTTGAGCTCCCAATTTGAGACAGTGTTCATCTTCG	899
Db	121	AGTGAATATACAAAAATATTCAAGGTGGAGCTCCCAATTTGAGACAGCTTCATCTTCG	180
QY	900	TTCTTCATATTACCCTGGCTGCCCAGAAAGCTTTGAGATTCAGATTGATACGAGTCA	959
Db	181	GCCTTCCTACTTACCACCTGGCTGTCCAGAAAGCTTTGAGATTCAGATTGATACGAGTCA	240
QY	960	TGTGATCTTGCAGTGTGGTGGGTATTCGCCAGTTTGTCAAGTACTTGATTCGCCCAAC	1019
Db	241	TGTGATCTTGCAGTGTGGTGGGTATTCGCCAGTTTGTCAAGTACTTGATTCGCCCAAC	300
QY	1020	CTGGCTGGAAAAAGAAATAGAAAGGCCACCAAGAAAGTACAGCTTCGCAATCCAGTTAT	1079
Db	301	TTGGCTAGAAAAAGAAATAGAAAGGCCACCAAGAAAGTACAGCTTCGCAATCCAGTAT	360
QY	1080	TGGAGTCCATTTTGACGCGACAGCAAAAGTGGGAGGGAAGAGGCTTCATCCCATTTGA	1139
Db	361	TGGAGTCCATTTTGACGCGACAGCAAAAGTGGGAGGGAAGAGGCTTCATCCCATTTGA	420
QY	1140	GGAATACACGGTGCACGCTTGAAGAGACTTTCAGCTTCTTGCCTGCAGATATCGAAGTGA	1199
Db	421	AGAGTACATGCTACATGTTGAAGAACATTTTTCAGCTTCTTGCACGAGATATCGAAGTGA	480
QY	1200	TAAAAAAAGGTGTAATTTGGCCACAGATGACCCTGTTTGTTTAAAGAGGCAAAAAACAA	1259
Db	481	TAAAAAAAGACTATATCTGCTACGCGATGACCTGCTTTGTTAAAGAGGCAAAAAACAA	540
QY	1260	GTACCCCGATTATGATTTTATAGTGTAACTCTATCTGTGGTCAGTGGACGTACATA	1319
Db	541	GTACTCCCATTTATGATTTTATAGTGTAACTCTATCTGTGGTCAGTGGACGTACATA	600
QY	1320	TGCAATATACAGAAATTCACCTCGGGGTGTGATCTGATATACACTTTCCTCCAGGC	1379
Db	601	TGCGTATACAGAAATTCACCTCGGGGGCGTGAATCGATATACACTTTCCTCTCAGGC	660
QY	1380	AGACTTCTAGTGTGTACTTTTTCATCGCAGGTCTGTAAAGTTGGCTTATGAAATCATGCA	1439
Db	661	TGACTTCTAGTGTGTACTTTTTCATCCAGGTCTGTGGGTGGCTTATGAAATCATGCA	720
QY	1440	AGCGTGCATCTGTATGCTCTGGGAACTTCGTTTGTGGATGACATCTACTATTTTGG	1499
Db	721	TAACCTGCATCTGTATGCTCTGCAAACTTCGACTTTTATGATGACATCTACTATTTTGG	780
QY	1500	AGGCCCAATGCCCAACAACCAAAATTTGCATTTATCTCACCAACCTCGAATGGAAGAGA	1559
Db	781	AGGCCCAATGCCCAACAACCAAGATTTGCCCTTATCTCACCAACCTCGAATGGAAGAGA	840
QY	1560	AATCCCATTCGAACCTGAGATATTATTTGTGTGGCTGGAAATTCACGTGGATGGCTATTC	1619
Db	841	AATTCATTCGAACCTGAGATATCATTTGTGTGGCTGGAAACCATTTGGAGTGTATTTC	900
QY	1620	TAAAGGCTTTAAGCAAAACGTGGGAAGCGAGGCTATATTCCTCCTACAAAGTTTCGAGA	1679
Db	901	TAAAGCTTCACACGAAAACCTTGGAAAACAGGCTTATATCTCCTCTACAAAGTCCGAGA	960
QY	1680	GAAATATGAAAACGTCAAG	1698

```
Db 961 GAAGATAGAAACGGTCAAG 979
|||||
RESULT 13
ABK70037
ID ABK70037 standard; cDNA; 979 BP.
XX
XX ABK70037;
AC
XX 15-JUL-2002 (first entry)
XX
XX Antibody production method related cDNA #5.
DE
XX
XX Antibody production; cytostatic; immunomodulator; vasotropic; virucide;
XX antibacterial; antiinflammatory; antiallergic; allergy; inflammation;
XX autoimmune disease; Chinese hamster ovarian tissue-originated cell; CHO;
XX tumor; circulatory disease; infection; primer: ss.
OS
XX Rattus norvegicus.
XX
XX WO200231140-A1.
XX
XX 18-APR-2002.
XX
XX 05-OCT-2001; 2001WO-JP08804.
XX
XX 06-OCT-2000; 2000JP-0308526.
XX
XX (KYOM ) KYOMA HARUKO KOGYO KK.
XX
XX Kanda Y, Satoh M, Nakamura K, Uchida K, Shinkawa T, Yamane N;
XX Hosaka E, Yamano K, Yamasaki M, Hanai N;
XX
XX WPI: 2002-340182/37.
XX
XX Cells producing antibody compositions including antibody fragments and
XX fusion proteins with Fc domain of antibody, useful for prevention or
XX treatment of cancer, immune diseases, circulatory diseases and
XX infections -
XX
XX Example 9: Page 13-14: 314pp; Japanese.
XX
XX This invention relates to novel method for antibody production
XX comprising a Chinese hamster ovarian tissue-originated (CHO) cell
XX transfected with a gene encoding an antibody molecule for producing a
XX composition comprising an antibody molecule with an Fc domain bonded
XX to the N-glycoside linkage complex sugar chain. The produced antibody
XX compositions are drugs for prevention or treatment of diseases
XX accompanying tumor, allergy or inflammation, autoimmune diseases,
XX circulatory diseases, and viral and bacterial infections. The
XX antibodies can be stably produced using the method of the invention
XX with high binding activity and potency thus leading to high safety and
XX reduced side effects when applied alone or in combination with other
XX drugs for therapy. The present sequence represents a nucleotide
XX molecule used in the method of the invention.
XX
XX Sequence 979 BP; 283 A; 223 C; 220 G; 253 T; 0 other;
XX
XX Query Match 47.9%; Score 827; DB 24; Length 979;
XX Best Local Similarity 90.3%; Pred. No. 2,9e-226;
XX Matches 884; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
```

```
Db 121 AGTGAATGACAAAATAATTCAGTGTGGAGCTCCCATTTGTAGACAGCCTTCATCTTCG 180
QY TCCTCATATTTTACCCCTGGCTGTCCAGAGACACCTTGCAGATTCGACTTGTACAGTCCA 959
Db 181 GCCTCCTTACTTACACACCTGGCTGTCCAGAGACACCTTGCAGATTCGACTTGTAGAGTCCA 240
QY TGGTGATCCTCGAGTGTGGTGTATCCAGTTCGATTCGATTCGATTCGATTCGATTCGATTCG 1019
Db 241 TGGTGATCCTCGAGTGTGGTGTATCCAGTTCGATTCGATTCGATTCGATTCGATTCGATTCG 300
QY CTGGCTGGAAGAAAGAAATGAGAGAGCCACCAAGAGCTTTCGATTCGATTCGATTCGATTCG 1079
Db 301 TGGTGATCCTCGAGTGTGGTGTATCCAGTTCGATTCGATTCGATTCGATTCGATTCGATTCG 360
QY 1080 TGGAGTTCATGTTAGACGACAGACAAAGTGGAGCGGAAGACGCTTCATCCATTCGA 1139
Db 361 TGGAGTTCATGTTAGACGACAGACAAAGTGGAGCGGAAGACGCTTCATCCATTCGA 420
QY 1140 GGAATACAGGAGTGGAGTGGAGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1199
Db 421 AGAGTACATGTGATGATGTTGAGAGACATTTTCAAGCTTCGACGAGAGTCAAGTGA 480
QY 1200 TAAAAAAGAGTGTATTTTGGCCACAGATGACCTGCTTTGTTAAAAAGAGCAAAACANA 1259
Db 481 TAAAAAAGAGTGTATTTTGGCCACAGATGACCTGCTTTGTTAAAAAGAGCAAAACANA 540
QY 1260 GTACCCAGTATGAAATTTATGATGATTAACCTCTCTTGGTCACTGACATACATA 1319
Db 541 GTACTCCAAATTTGAAATTTATGATGATTAACCTCTCTTGGTCACTGACATACATA 600
QY 1320 TCGATTCACAGAAATTCATCTTGGGGGTGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 1379
Db 601 TCGATTCACAGAAATTCATCTTGGGGGTGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 660
QY 1380 AGACTTCCTAGTGTGATCTTTTTCATCGCAGGTGTGATGATGATGATGATGATGATGATGATG 1439
Db 661 TGACTTCCTAGTGTGATCTTTTTCATCGCAGGTGTGATGATGATGATGATGATGATGATGATG 720
QY 1440 AGCGTGCATCTGATGCTCTGCGACATTCCTTTTGGATGATGATGATGATGATGATGATGATG 1499
Db 721 AACCCGTGATCTGATGCTCTGCGACATTCCTTTTGGATGATGATGATGATGATGATGATGATG 780
QY 1500 AGGCCCAAAATGGCCACACCAAAATGGCCATTTATCTCTCAACCTCGAAGTGAAGAGA 1559
Db 781 AGGCCCAAAATGGCCACACCAAAATGGCCATTTATCTCTCAACCTCGAAGTGAAGAGA 840
QY 1560 AATCCCATGGAACCTGAGATATTTTGGTGTGCTGGAATTCATCGATGATGATGATGATGATGATG 1619
Db 841 AATCCCATGGAACCTGAGATATTTTGGTGTGCTGGAATTCATCGATGATGATGATGATGATGATG 900
QY 1620 TAAAGGTGTTAACAGAAAACCTGGGAAGAGCGGCTTATTCCTCTACAAAGTTCGAGA 1679
Db 901 TAAAGGTGTTAACAGAAAACCTGGGAAGAGCGGCTTATTCCTCTACAAAGTTCGAGA 960
QY 1680 GAAGATAGAAACAGTCAAG 1698
Db 961 GAAGATAGAAACAGTCAAG 979
XX
XX RESULT 14
XX AAF87953
XX ID AAF87953 standard; DNA; 699 BP.
XX
XX AAF87953;
XX
XX 20-JUL-2001 (first entry)
XX
XX Human alpha 1-6 fucosyltransferase 343-575 encoding DNA SEQ ID NO:2.
XX
XX Human; alpha 1-6 fucosyltransferase; alpha 1-6 fucT; antibody;
XX alpha 1-6 fucose transferase; anti-human alpha 1-6 fucT; immunosay;
XX ds.
```



```

OS Homo sapiens.
XX Key Location/Qualifiers
FH CDS 1..689
FT /tag=
FT /partial=
FT /product= "alpha 1-6 fucosyltransferase 343-575"
PN JP2001011097-A.
PD 16-JAN-2001.
PE 29-JUN-1999. 99JP-0183569.
PR 29-JUN-1999. 99JP-0183569.
XX (FJRE ) FUJIREBIO KK.
PA WPI: 2001-275926/29.
DR P-PSDB: AAB75062.
PT Novel anti-human alpha1-6 fucose transferase antibody useful for
PT immunosassay -
PS Example 1; Page 7-8; 11pp; Japanese.
XX
XX The present invention describes an anti-human alpha 1-6
XX fucosyltransferase (alpha 1-6 fuct, also called alpha 1-6 fucose
XX transferase) antibody. Also described are: (1) a hybridoma producing
XX the above monoclonal antibody; (2) an immunosassay for detecting human
XX alpha 1-6 fuct by using the above antibody or its antibody fragment;
XX and (3) a reagent used for the above immunosassay. The anti-human
XX alpha 1-6 fuct antibody can be used for immunosassay. The present
XX sequence encodes the human alpha 1-6 fuct protein of residues 343 to
XX 575, which is used in an example from the present invention.
SQ Sequence 699 BP; 220 A; 152 C; 147 G; 180 T; 0 other;

Query Match 35.1%; Score 606.8; DB 22; Length 699;
Best Local Similarity 91.8%; Pred. No. 2.8e-163;
Matches 641; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

OY 1027 GAAAGGAAATAGAGAGGCGACCAAGAGCTAGGCTTCAACATCAGTTATTGAGTC 1086
DB 1 GAAAGGAAATAGAGAGGCGACCAAGAGCTTCAACATCAGTTATTGAGTC 60
OY 1087 CATGTAGACGCGACAGCAAGAGTGGAGCGAGCAGCCTTCATCCATGAGGAATAC 1146
DB 61 CATGTAGACGCGACAGCAAGAGTGGAGCGAGCCTTCATCCATGAGGAATAC 120
OY 1147 ACGGTGACAGTGGAGAGAGCTTTCAGCTTCTGTCGAGAAATGCAAGTGATAAAAA 1206
DB 121 ATGGTGCATGTTGAAGAACTTTTCAAGCTTCTTGCACGCGAATGCAAGTGACAAAAA 180
OY 1207 AGGGTGTATTTGGCCACAGATGACCTGCTTGTAAAGAGCAAAAGAGTACCCC 1266
DB 181 AGAGTGTATTTGGCCACAGATGACCTTGTATTAAGAGGCAAAAGAGTACCCC 240
OY 1267 AGTTATGAAATTTATGATATCTATCTCTGTCAGCTGAGCTGAGTACTATATCATAT 1326
DB 241 AATTTAATTAATTTATGATATCTATCTCTGTCAGCTGAGTACTATATCATAT 300
OY 1327 ACAGAAATTTCACTTGGGGTGTATCTGATATACATTTCTCTCCAGGAGACTTC 1386
DB 301 ACAGAAATTTCACTTGGGGTGTATCTGATATACATTTCTCTCTCAGGAGACTTC 360
OY 1387 CTAGTGTGACTTTTTCATCGCAGGCTGTAGAGTGCCTTATGAATCATCAAGGCTG 1446
DB 361 CTAGTGTGACTTTTTCATCGCAGGCTGTAGAGTGCCTTATGAATCATCAAGACTA 420
OY 1447 CATCCGATGCGCTCTCGCAACTTCGCTTCTTGGATGACATCTACTATTTTGAAGCCCA 1506
DB 421 CATCCGATGCGCTCTCGCAACTTCGCTTCTTGGATGACATCTACTATTTTGGGCGCAG 480

```

```

OY 1507 AATGCCACAAACCAATTTGCAATTTATCTCACCACCTGACACTGAGAGAAATCCCC 1566
DB 481 AATGCCACAAATTAATTTGCAATTTATGCTCACCAACCCCGAACTGACATGAAATTTCC 540
OY 1567 ATGGAACCTTGAGATATTTATTTGTTGCTGCTGGAATCACTGAGGAGCTATCTTAAGCT 1626
DB 541 ATGGAACCTTGAGATATTTATTTGTTGCTGCTGGAATCACTGAGGAGCTATCTTAAGCT 600
OY 1627 GTTACAGAGAAACTGGGAGGAGGAGGCTATATCCCTCCACAAAGTTGAGAGAGATA 1686
DB 601 GTTACAGAGAAATTTGGAGAGGAGGCTATATCCCTCCACAAAGTTGAGAGAGATA 660
OY 1687 GAAACAGTCAAGTATCCACATATCCCGAGGCTGAGAA 1724
DB 661 GAAACGTTCAAGTATCCCGACATATCCGAGGCTGAGAA 698

RESULT 15
ABLO4601
ID ABLO4601 standard; cDNA; 2761 BP.
XX
XX AC ABLO4601;
XX
XX DT 26-MAR-2002 (first entry)
XX
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 8285.
XX
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX
XX PD 27-SEP-2001.
XX
XX PF 23-MAR-2001; 2001WO-US09231.
XX
XX PR 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX
XX PA (PEKE ) PE CORP NY.
XX
XX PI Venter JC, Adams M, Li PMD, Myers EW;
XX
XX DR WPI: 2001-656860/75.
XX P-PSDB: ABB60498.
XX
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX PS Claim 1; SEQ ID NO 8285; 21pp + Sequence Listing; English.
XX
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABBS7737-ABBS7072).
XX
XX CC The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2761 BP; 657 A; 731 C; 755 G; 618 T; 0 other;

Query Match 20.3%; Score 351.6; DB 23; Length 2761;
Best Local Similarity 55.9%; Pred. No. 6.9e-90;
Matches 694; Conservative 0; Mismatches 539; Indels 9; Gaps 1;

```


OY	473	ATCATGAAAGGCTATTAATAGAGGATCTTACTACTAGTCAAAAGATGGGCGAGTGC	552
OY	1083	ACGCAAGCGCTCCCTGCTGACGATATGAGCGGATGCGTCACTGCTCGATGGCTTACGAGG	1142
OY	533	ATTGGCTGAAAGAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGGCAATTAACATATC	592
OY	1143	CCTGCGACACAGAGGAGGCAJCGGATTTTGAAGCATTTTGGTACAGCGCGCCCTGCACCATC	1202
OY	593	TTTCAATCCCAAGGAGCTGCAGCAAAAGCCAGAAGCTAGTGTATTAATCAACAAAGCT	652
OY	1203	TACAGAAATCCGAGTATTCAGAAATGCTCGCAACCTGGTCTGTAGCTCAACAAAGGCT	1262
OY	653	GTGGCTATGGGCTGACGCTCCATCATGTAGTACTGCTTATGATTTGCATATGGCACCC	712
OY	1263	GTGGCTATGGGCTGACGCTCCATCATGTAGTACTGCTTATGATTTGCATATGGCACCC	1322
OY	713	AGCGAAGCTGCGCTTGGAAATCTCACAATTTGGCGCTACCTACTGGGGATGGGAAACTG	772
OY	1323	AGCGAAGCTGCGCTTGGAAATCTCACAATTTGGCGCTACCTACTGGGGATGGGAAACTG	1382
OY	773	TGTTAGACCTGTAAAGTGAAGCTGACAGACAGATCTGGCAGCTCCACTGACATTTGGT	832
OY	1383	TGTTAGACCTGTAAAGTGAAGCTGACAGACAGATCTGGCAGCTCCACTGACATTTGGT	1442
OY	833	CAGGTGAAGTAAAGACAAAAATGTTCAAGTGGTTAGCTCCCATTTGATAGACGTGTC	892
OY	1443	CAGGTGAAGTAAAGACAAAAATGTTCAAGTGGTTAGCTCCCATTTGATAGACGTGTC	1493
OY	893	ATCTCTGCTCCATATTTAACCCCTGGCTGTCCAGAAAGCCTTGAGATGCACTTGTAC	952
OY	1494	TGCCAGAGACCGCGTACCTGCGCTGCTGCGCTTCCGAGATCTGGCGCGGAGCTTAAGC	1553
OY	953	GAGTCATGCTGATCTGCAATGTGGTGGTATCCAGTTTGTCAAGTACTTGAATTCGCC	1012
OY	1554	GTCGATGATGAGAACCCCATCTGTTGGTGGTGGTCACTTCTTAAGTATTTGCTCCGAC	1613
OY	1013	CACAAACCTGGCTGGAAGAAAGAAATGAAGAGGCCACCAAGAAAGCTTCAAAATTC	1072
OY	1614	CACAAACCTGGCTGGAAGAAAGAAATGAAGAGGCCACCAAGAAAGCTTCAAAATTC	1673
OY	1073	CAGTTATGGATTCATGTATAGCGACAGACAAAGTGGAGCGGAGAGCAGCTTCATC	1132
OY	1674	CCATTGTTGGCTTCTGCTGCTGCGACGACAAAGTGGGACACAGAGCGGCTGCGACA	1733
OY	1133	CCATTGAGGAATACAGGCTGACGTTGAAGAAAGCTTTCAGCTTCTGCTCCGAGATGC	1192
OY	1734	GTCGTGAGGAGATATATGACCTATGTGAGAGACTACTACCGCACGCTGAGGTAAAGCGCA	1793
OY	1193	AAAGTGAATAAAGAGGTGTATTTGGCCACAGATGACCCCTGCTTGTAAAGAGGCAA	1252
OY	1794	GCACGTGGCTGCGCGGATTTCTTCGCAATCGGATGATCCCGAGTTATTAAGAGGCGC	1853
OY	1253	AAAGTGAATAAAGAGGTGTATTTGGCCACAGATGACCCCTGCTTGTAAAGAGGCAA	1312
OY	1854	GCCGAAAGTACCCCGAGTACAGATCATTTGGTATCCGAGGCTGGCGCGCATGGCGCTCG	1913
OY	1313	TACATATGATATACAGAAATTCACCTCGGGGTGTGATCTCGATATACACTTCTCT	1372
OY	1914	TATCACAAGATACACAGGACACCGCCCTGAACGGGATCATTTCTGATATCCACTGCTTT	1973
OY	1373	CCCAGGACAGACTTCTAGTGTACTTTTTCATTCGAGGCTGTGTAGAGTTGCTTATGAA	1432
OY	1974	CCATGTCGATCATGTGGTGTGACCTTCTGTCGAGGTGTGTGCGTGGCTTACGAGA	2033
OY	1433	TCATGCAAGCGTGCATCCTGATGCTCTGCGAAGCTTCGCTTCTTGGATGACATCTACT	1492
OY	2034	TATATCAGACGATGATTCGATGCAAGCGCATCGGTTCAAGTGTGAGAGACATATATCT	2093
OY	1493	ATTTTGAAGGCCCAATGCGCACAAATTTGCCATTTATCCTCACCAACCTCGAACTG	1552
OY	2094	ACTAGGTTGGCAGATGGGACAAATGCGCGCTGCTTATCGCCCAAGCGCGCACGCG	2153
OY	1553	AAGGAAATCCCATGGAACCTGAGATATTAATGGTGTGCTGGGAAATCACTGGGATG	1612

Search completed: November 1, 2002, 21:39:33
Job time : 286 secs

OY	1613	GCTATCCTAAAGGTGTATACGAAAACTGGGAAAGACGGGCTTATTCCTCTACAAAG	1672
OY	2214	GCAATTTCAAGGGGCAAGAAATACCCGACACCAACGAGGGCGGCTTATTCCTCTCTCAAG	2273
OY	1673	TTGAGAGAGATTAAGAAACAGTCAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT	1714
OY	2274	TGAGAGAGAGGTGACACCGCCAAAGCTGCTCTATGCGG	2315

Search completed: November 1, 2002, 21:39:33
Job time : 286 secs

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2002, 20:37:08 : Search time 3099 Seconds

(without alignments)
16227.702 Million cell updates/sec

Title: US-09-839-136-1

Perfect score: 1728
Sequence: 1 atggcgccatgactgcttc.....atcccgagctgacagtaaa 1728

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_srs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_srs:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rpd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1726.4	99.9	1728	6	AR170074	AR170074 Sequence
2	1726.4	99.9	1728	6	E14720	E14720 Proline mRN
3	1726.4	99.9	1836	4	D86723	D86723 Porcine mRN
4	1545.6	89.4	1728	4	AF247186	AF247186 Bos tauru
5	1510.4	87.4	2002	9	D89289	D89289 Homo sapien
6	1510.4	87.4	2100	6	E15725	E15725 Human mRN
7	1510.4	87.4	2898	9	HSY17976	HSY17976 Homo sapien
8	1510.4	87.4	2992	9	HSY17977	HSY17977 Homo sapien
9	1510.4	87.4	3186	9	HSY17978	HSY17978 Homo sapien
10	1510.4	87.4	3280	9	HSY17979	HSY17979 Homo sapien
11	1505.6	87.1	2100	6	AR170077	AR170077 Sequence
12	1467.8	84.9	2976	10	BC010666	BC010666 Mus muscu
13	1464.6	84.8	2176	10	AB025198	AB025198 Mus muscu
14	1215.4	70.3	2682	9	BC025385	BC025385 Homo sapi
15	950	55.0	1850	9	AF052088	AF052088 Homo sapi
16	878.4	50.8	1017	6	E43925	E43925 Antlbody re
17	750	43.4	1772	9	AB049740	AB049740 Homo sapi
18	606.8	35.1	699	6	E43926	E43926 Antlbody re
19	351.6	20.3	2764	3	AY051451	AY051451 Drosophil
20	276.8	16.0	1147	9	AB049828	AB049828 Homo sapi
21	276.8	15.7	186158	2	CNS018P0	AL109847 Human chr
22	270.8	15.7	187192	2	AC128532	AC128532 Rattus no
23	219.4	12.7	359	9	HSALSFUC1	AF038280 Homo sapi
24	215	12.4	34322	2	AC018043	AC018043 Drosophil
25	215	12.4	160038	2	AC023739	AC023739 Drosophil
26	213	10.3	366	9	AB032576	AB032576 Homo sapi
27	185.4	10.7	148068	9	CNS018RN	AL161871 Human chr
28	183.8	10.6	662	9	AB03256752	AB032568 Homo sapi
29	175.8	10.2	168601	10	AC110302	AC110302 Mus muscu
30	155.4	9.0	335	9	HSALSFUC2	AF038281 Homo sapi
31	149.6	8.7	407	9	AB03256754	AB032570 Homo sapi
32	138.6	8.0	187192	2	AC128532	AC128532 Rattus no
33	136.6	7.9	351	9	AB03256757	AB032573 Homo sapi
34	129.2	7.5	221521	2	AC129041	AC129041 Rattus no
35	108.6	6.3	75282	2	DMBR37M9	AL133495 Drosophil
36	107.8	6.2	125	9	AB03256755	AB032571 Homo sapi
37	98	5.7	40172	3	AF022968	AF022968 Caenorhab
38	95.4	5.5	264	9	AB03256753	AB032569 Homo sapi
39	75.6	4.4	7218	6	I66494	I66494 Sequence 14
40	72.4	4.2	68336	2	AC100653	AC100653 Mus muscu
41	50	2.9	164304	2	AP002398	AP002398 Homo sapi
42	50	2.9	174569	2	AC011123	AC011123 Homo sapi
43	50	2.9	208532	2	AP001460	AP001460 Homo sapi
44	50	2.9	210839	2	AP001562	AP001562 Homo sapi
45	50	2.9	215374	2	AC090396	AC090396 Homo sapi

ALIGNMENTS

RESULT 1
AR170074 LOCUS
DEFINITION Sequence 1 from patent US 6291219.
ACCESSION AR170074
VERSION AR170074.1 GI:17908033
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1728)
AUTHORS Taniguchi,N., Uozumi,N., Shiba,T. and Yanagida,I.S.
TITLE .alpha.1-6 fucosyltransferase
JOURNAL Patent: US 6291219-A 1 18-SEP-2001;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

source 1..1728 /organism="unknown"
BASE COUNT 521 a 362 c 419 g 426 t
ORIGIN
Query Match 99.9%: Score 1726.4: DB 6: Length 1728:
Best Local Similarity 99.9%: Pred. No. 0:
Matches 1727: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
1 ATGCGGCATGACGCTGTTGCGGTGGATGATGCTCATCTCTTTTGGCTGGGACC 60
1 ATGCGGCATGACGCTGTTGCGGTGGATGATGCTCATCTCTTTTGGCTGGGACC 60
1 ATGCGGCATGACGCTGTTGCGGTGGATGATGCTCATCTCTTTTGGCTGGGACC 60
61 TTGCTATTTTACATAGAGTGGTCACTTGGTACGAGATATGACCACTGTGACTCTAGC 120
61 TTGCTATTTTACATAGAGTGGTCACTTGGTACGAGATATGACCACTGTGACTCTAGC 120
121 CGAAGACTGTCAGATTTTGGCAAGCTGGAAAGCTTAAACACAAATGAAGACTTG 180
121 CGAAGACTGTCAGATTTTGGCAAGCTGGAAAGCTTAAACACAAATGAAGACTTG 180
181 AGAGAGATGGCTGATCTCTCGAATACAGAAAGGCCCATTTGATGAGGGCCAGCTTCA 240
181 AGAGAGATGGCTGATCTCTCGAATACAGAAAGGCCCATTTGATGAGGGCCAGCTTCA 240
241 GGAAGATTCGTCCTTTAGAAAGCAATTTATGAGGCCAAAGACAGATTTGAAATTTAT 300
241 GGAAGATTCGTCCTTTAGAAAGCAATTTATGAGGCCAAAGACAGATTTGAAATTTAT 300
301 AAGAAACAACTAAAATGGTCCAGGAAAGATCATGAATCTTAAAGAGAGATTTGAA 360
301 AAGAAACAACTAAAATGGTCCAGGAAAGATCATGAATCTTAAAGAGAGATTTGAA 360
361 AATGAGACTAAAGAGCTGTTTCTTCTACAAAGTGAAGTGAAGAAATTAAGAAATTTA 420
361 AATGAGACTAAAGAGCTGTTTCTTCTACAAAGTGAAGTGAAGAAATTAAGAAATTTA 420
421 GAAGAAATGAACTCCAAAGACATGAGATGAAATTTCTATCAGATTTGGGACATGAA 480
421 GAAGAAATGAACTCCAAAGACATGAGATGAAATTTCTATCAGATTTGGGACATGAA 480
481 AGGTCTATATGACGATCTATCTACTCTGATGCAAAAGATGGGACGATTTGGCGT 540
481 AGGTCTATATGACGATCTATCTACTCTGATGCAAAAGATGGGACGATTTGGCGT 540
541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGGAGGAATTAACATATCTTCAGAAAT 600
541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCAGGAGGAATTAACATATCTTCAGAAAT 600
601 CCCAAGGACTGCAGCAAAAGCCAAAGCTAGTGTATATCAACAAGGCTGTGGCTAT 660
601 CCCAAGGACTGCAGCAAAAGCCAAAGCTAGTGTATATCAACAAGGCTGTGGCTAT 660
661 GGTGTACGATCATCATGATGATGATGCTGTTATGATGATGATGATGATGATGATGAT 720
661 GGTGTACGATCATCATGATGATGATGCTGTTATGATGATGATGATGATGATGATGAT 720
721 CTGCGCTTGAATCTCACAATTTGGCGCTACGCTACTGAGGAGATGGGAACTGTGTTAGA 780
721 CTGCGCTTGAATCTCACAATTTGGCGCTACGCTACTGAGGAGATGGGAACTGTGTTAGA 780
781 CCTGTAAAGTGAAGCTGACAGACAGATCTGAGCTCCACTGAGATTTGGTCAAGTGA 840
781 CCTGTAAAGTGAAGCTGACAGACAGATCTGAGCTCCACTGAGATTTGGTCAAGTGA 840
841 GTTAAAGGACAAAAATGTTACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 900
841 GTTAAAGGACAAAAATGTTACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 900
901 CCTCAATATTTACCCCTGGCTGTCCAGAGAGCTTGCAGATGACCTTGTACAGATTCAT 960
901 CCTCAATATTTACCCCTGGCTGTCCAGAGAGCTTGCAGATGACCTTGTACAGATTCAT 960

QY 961 GGTGATCTGCACTGTGTGGGATATCCAGATTTGTCAAGTACTTGAATTCGCCACACACC 1020
DB 961 GGTGATCTGCACTGTGTGGGATATCCAGATTTGTCAAGTACTTGAATTCGCCACACACC 1020
QY 1021 TGGCTGGAAGAAATGAGAGGCGCACCAAGAACTAGGCTTCAAACTACCTATTT 1080
DB 1021 TGGCTGGAAGAAATGAGAGGCGCACCAAGAACTAGGCTTCAAACTACCTATTT 1080
QY 1081 GGAATCATGTTTACAGCAGACAGCAAAAGTGGAGCGGAAGACCTTCCATCCATTTAG 1140
DB 1081 GGAATCATGTTTACAGCAGACAGCAAAAGTGGAGCGGAAGACCTTCCATCCATTTAG 1140
QY 1141 GAATACAGGTCACCTTGAAGAGACTTTCAGCTTCTTCTGCGCAATTCGAAGTGAAT 1200
DB 1141 GAATACAGGTCACCTTGAAGAGACTTTCAGCTTCTTCTGCGCAATTCGAAGTGAAT 1200
QY 1201 AAAAAAGGCTATTTGGCCACAGATGACCCGCTTGTAAAGAGCAAAACAAAG 1260
DB 1201 AAAAAAGGCTATTTGGCCACAGATGACCCGCTTGTAAAGAGCAAAACAAAG 1260
QY 1261 TACCCAGTTATGATTTATTTAGTAACTATCTCTTGGTCAAGTCACTAAT 1320
DB 1261 TACCCAGTTATGATTTATTTAGTAACTATCTCTTGGTCAAGTCACTAAT 1320
QY 1321 CGATATACAGAAATTTACCTCGGGGTGTGATCTGATATACACTTCTCCAGGCA 1380
DB 1321 CGATATACAGAAATTTACCTCGGGGTGTGATCTGATATACACTTCTCCAGGCA 1380
QY 1381 GACTTCTAGTGTGATTTTTCATGCGAGCTCTGATGAGTGTGATGAGTGTGATGAGT 1440
DB 1381 GACTTCTAGTGTGATTTTTCATGCGAGCTCTGATGAGTGTGATGAGTGTGATGAGT 1440
QY 1441 GCGCTCATCTGATGATCTCTGCGAATCTCCGCTTCTTGGATGACATCTATTTTGA 1500
DB 1441 GCGCTCATCTGATGATCTCTGCGAATCTCCGCTTCTTGGATGACATCTATTTTGA 1500
QY 1501 GGGCCAAATGCCACAAACAAATTTGCCATTTATCTTCCACCACTGGAAGAGAA 1560
DB 1501 GGGCCAAATGCCACAAACAAATTTGCCATTTATCTTCCACCACTGGAAGAGAA 1560
QY 1561 ATCCCATGGAACCTGAGATATTTTGGTGTGGGAAATCAGGGGATATCTT 1620
DB 1561 ATCCCATGGAACCTGAGATATTTTGGTGTGGGAAATCAGGGGATATCTT 1620
QY 1621 AAAGGTGTTAACGAAACCTGGGAAGAGGAGGCTTATTCCTCTTCAACAAGTTCAGAG 1680
DB 1621 AAAGGTGTTAACGAAACCTGGGAAGAGGAGGCTTATTCCTCTTCAACAAGTTCAGAG 1680
QY 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAGTAA 1728
DB 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGACAGTAA 1728
RESULT 2
E14720 1728 bp DNA linear PAT 28-JUL-1999
LOCUS E14720
DEFINITION Procine mRNA for alpha-1,6-fucosyltransferase.
ACCESSION E14720.1 GI:5709403
VERSION E14720.1
KEYWORDS JP 1998004969-A/1.
SOURCE Sus sp.
ORGANISM Sus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 1728)
AUTHORS Taniguchi, N., Uozumi, H. and Shiba, T.
TITLE ALPHA-1,6-FUCOSYLTRANSFERASE GENE
JOURNAL Patent: JP 1998004969-A 1 13-JAN-1998;
TOYOBO CO LTD
COMMENT OS Sus sp. (p19)
PN JP 1998004969-A/1
PD 13-JAN-1998
PF 24-JUN-1996 JP 1996162813

PI TANICUCHI NAOKUKI, UOZUMI HISAFUMI, SHIBA TETSUO PC
C12N15/09.C07H21/04.C12N5/10.C12N9/10.(C12N5/10.C12R1.91). PC
(C12N9/10.
PC (C12R1.91):
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FT source 1..1728
FT /organism='Sus sp.'
FT /tissue_type='brain'
FT CDS 1..1728
FT /product='alpha-1,6-fucosyltransferase'
FEATURES
source location/Qualifiers
1..1728
/organism='Sus sp.'
/db_xref='taxon:9826'
BASE COUNT 521 a 362 c 419 g 426 t
ORIGIN
Query Match 99.9%; Score 1726.4; DB 6; Length 1728;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ATGCGGCATGACATGCTGCTGCGTGGATTAATGCTCATCTTTTGGCTGGGGACC 60
DB 1 ATGCGGCATGACATGCTGCTGCGTGGATTAATGCTCATCTTTTGGCTGGGGACC 60
OY 61 TTGCTATTTTACATAGGTGCTCACTTGGTACGAGATTAATGACCACTCTGATCACTAGC 120
DB 61 TTGCTATTTTACATAGGTGCTCACTTGGTACGAGATTAATGACCACTCTGATCACTAGC 120
OY 121 CGGAACTGTCGCAAGATTTTGGCAAGCTGGAAACCGCTTAACAACAATAATGAAGCTTG 180
DB 121 CGGAACTGTCGCAAGATTTTGGCAAGCTGGAAACCGCTTAACAACAATAATGAAGCTTG 180
OY 181 AGAGAAATGGCTGATCTCTCCGATACCGAAGGCCCTTGTATGAGGGGCCAGCTTCA 240
DB 181 AGAGAAATGGCTGATCTCTCCGATACCGAAGGCCCTTGTATGAGGGGCCAGCTTCA 240
OY 241 GGAAGATTGCTGCTTTAGAGAGCAATTTATGAAGGCCCAAGACAGATTTGAATAATTA 300
DB 241 GGAAGATTGCTGCTTTAGAGAGCAATTTATGAAGGCCCAAGACAGATTTGAATAATTA 300
OY 301 AAGAAACAACTAAATATGTCGAGGGAAGATCATGAATTCCTAAGGAGAGAGATTGAA 360
DB 301 AAGAAACAACTAAATATGTCGAGGGAAGATCATGAATTCCTAAGGAGAGAGATTGAA 360
OY 361 AATGAGCTAAAGAGCTCTGTTTTTTCTCAAGAGTGAATTTGAAGAAATTTAAAGATTTA 420
DB 361 AATGAGCTAAAGAGCTCTGTTTTTTCTCAAGAGTGAATTTGAAGAAATTTAAAGATTTA 420
OY 421 GAAGAAATGAACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGGACATCATGAA 480
DB 421 GAAGAAATGAACTCCAAAGACATGAGATGAATTTCTATCAGATTTGGGACATCATGAA 480
OY 481 AGGTCTATATGACGATCTATCTACTGATGTCAGTCAAGCAATGGGGAGGTGATGGCGT 540
DB 481 AGGTCTATATGACGATCTATCTACTGATGTCAGTCAAGCAATGGGGAGGTGATGGCGT 540
OY 541 GAAAGAGAGCCCAAGATCTGACAGAGCTGTCAGGCGAGATTAACATATCTTCAGAAAT 600
DB 541 GAAAGAGAGCCCAAGATCTGACAGAGCTGTCAGGCGAGATTAACATATCTTCAGAAAT 600
OY 601 CCCAAGAGCTGACAGCAAGCAAGCAAGCTAGTGTATATCAACAAGGCTGTGGCTAT 660
DB 601 CCCAAGAGCTGACAGCAAGCAAGCAAGCTAGTGTATATCAACAAGGCTGTGGCTAT 660
OY 661 GCGTGTACAGCTCATCATGTAGTGTACTGCTTATATGATTCATATGCGACCGCAGACA 720
DB 661 GCGTGTACAGCTCATCATGTAGTGTACTGCTTATATGATTCATATGCGACCGCAGACA 720
OY 721 CTGGCTTGAATCTCAATTTGGCGCTAGCTAGTGGGGATGGGAAACTGTGTTTAGA 780

DB 721 CTGGCTTGAATCTCAATTTGGCGCTAGCTAGTGGGGATGGGAAACTGTGTTTAGA 780
OY 781 CCTGTAGTGAAGACGTCACAGACAGATCTGGCAGCTCCACTGGACATTTGGCTAGTGA 840
DB 781 CCTGTAGTGAAGACGTCACAGACAGATCTGGCAGCTCCACTGGACATTTGGCTAGTGA 840
OY 841 GTTAAAGACAAAATGTTTACAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCCTCGT 900
DB 841 GTTAAAGACAAAATGTTTACAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCCTCGT 900
OY 901 CTTCAATTTTACCCCGGCTGTCCAGAAAGACCTTGCAGATGACATTTGAGAGTTCAT 960
DB 901 CTTCAATTTTACCCCGGCTGTCCAGAAAGACCTTGCAGATGACATTTGAGAGTTCAT 960
OY 961 GGTGATCTGACAGTGTGTTGAGTATCCCATTTGTCAAGTACTTGTATGCGCCCAACCC 1020
DB 961 GGTGATCTGACAGTGTGTTGAGTATCCCATTTGTCAAGTACTTGTATGCGCCCAACCC 1020
OY 1021 TGGCTGAAAAGAAATATGAAAGAGCGCACCAAGACCTAGGCTTCAAACTCCAGTTATT 1080
DB 1021 TGGCTGAAAAGAAATATGAAAGAGCGCACCAAGACCTAGGCTTCAAACTCCAGTTATT 1080
OY 1081 GGAATCCATGTTAGACGACAGCAAAAGTGGAGCGGAACGACCTTCCATCCATTGAG 1140
DB 1081 GGAATCCATGTTAGACGACAGCAAAAGTGGAGCGGAACGACCTTCCATCCATTGAG 1140
OY 1141 GAATACACGGTGCAGCTTGAAGAACTTTCACCTTCTGTCGCAAGATGCAAGTGAT 1200
DB 1141 GAATACACGGTGCAGCTTGAAGAACTTTCACCTTCTGTCGCAAGATGCAAGTGAT 1200
OY 1201 AAAAAAGGTTATTTTGGCCACAGATGACCTTGTGTTAAAGAGGCAAAACAAAG 1260
DB 1201 AAAAAAGGTTATTTTGGCCACAGATGACCTTGTGTTAAAGAGGCAAAACAAAG 1260
OY 1261 TACCCAGTATGAATTTATGATTAATCTATCTCTGTCGACCTGACATCAATTAAT 1320
DB 1261 TACCCAGTATGAATTTATGATTAATCTATCTCTGTCGACCTGACATCAATTAAT 1320
OY 1321 CGATATACAGAAATTTCACTTGGGGGTGATCCTGGATTTACACTTCTCTCCAGGCA 1380
DB 1321 CGATATACAGAAATTTCACTTGGGGGTGATCCTGGATTTACACTTCTCTCCAGGCA 1380
OY 1381 GACTTCCTAGTGTACTTTTTCATGCGAGTCTGTAAGATTTCTTAATGAATCATGCA 1440
DB 1381 GACTTCCTAGTGTACTTTTTCATGCGAGTCTGTAAGATTTCTTAATGAATCATGCA 1440
OY 1441 GCGCTGATCTGATGCTCTGCGAACTTCCGTTCTTTGGATGACATCTAATTTTGA 1500
DB 1441 GCGCTGATCTGATGCTCTGCGAACTTCCGTTCTTTGGATGACATCTAATTTTGA 1500
OY 1501 GGCCTCAATGCCCCAACAATTTGCCATTTATCTCACCACTCGAACTGAAGAGAA 1560
DB 1501 GGCCTCAATGCCCCAACAATTTGCCATTTATCTCACCACTCGAACTGAAGAGAA 1560
OY 1561 ATCCCATGGAACCTGAGATTAATTTGATGTTGATGGAATACAGTGGCTATCTCT 1620
DB 1561 ATCCCATGGAACCTGAGATTAATTTGATGTTGATGGAATACAGTGGCTATCTCT 1620
OY 1621 AAAGGTGTTAAACAGAAACCTGGAAGAGCGGCTATATCTCTCTCAAAAGTTCCAGAG 1680
DB 1621 AAAGGTGTTAAACAGAAACCTGGAAGAGCGGCTATATCTCTCTCAAAAGTTCCAGAG 1680
OY 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTCACAAGTAA 1728
DB 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTCACAAGTAA 1728
RESULT 3 1836 bp mRNA linear MAM 07-FEB-1999
D86723 Porcine mRNA for N-acetyl-beta-D-glucosaminide
DEFINITION alpha-1,6-fucosyltransferase, complete cds.

ACCESSION D86723
VERSION D86723.1 GI:1752753
KEYWORDS N-acetyl-beta-D-glucosaminide alpha-1,6-fucosyltransferase;
GDP-L-Fuc
SOURCE Sus scrofa adult brain CDNA to mRNA.
ORGANISM Sus scrofa
REFERENCE Mamalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
AUTHORS 1 (sites)
Uozumi, N., Yanagidani, S., Miyoshi, E., Ihara, Y., Sakuma, T.,
Gao, C.X., Teshima, T., Fujii, S., Shiba, T. and Taniguchi, N.
TITLE Purification and cDNA cloning of porcine brain
GDP-L-Fuc:N-acetyl-beta-D-glucosaminide
alpha-1->6fucosyltransferase
J. Biol. Chem. 271 (44), 27810-27817 (1996)
JOURNAL 97066976
MEDLINE
REFERENCE 2 (bases 1 to 1836)
AUTHORS Naoi, U., Shusaku, Y., Elj, M., Yoshito, I., Takahiko, S.,
Cong-Xiao, G., Tadashi, T., Shigeru, F., Tetsuo, S. and Naoyuki, T.
TITLE Purification and cDNA cloning of porcine brain
GDP-L-Fuc:N-acetyl-b-D-glucosaminide:alpha-1-6fucosyltransferase
Unpublished
JOURNAL 3 (bases 1 to 1836)
REFERENCE Uozumi, N.
AUTHORS Direct Submission
TITLE Submitted (29-JUL-1996) Naoi, Uozumi, Osaka University Medical
Journal School, Department of Biochemistry: Yamadaoka 2-2, Suita, Osaka
565, Japan (E-mail: seika@biochem.med.osaka-u.ac.jp,
tel:06-879-3421, fax:06-879-3423)
FEATURES
source
1. 1836
/organism="Sus scrofa"
/db_xref="taxon:9823"
/lisse.type="brain"
/dev_stage="adult"
57. 1784
/note="GDP-L-Fuc"
/product="N-acetyl-beta-D-glucosaminide
alpha-1,6-fucosyltransferase"
/protein_id="BA113157.1"
/db_xref="GI:1752754"
/translation="MRPWTGSNRMIMLILFANGTLLFYIGHLVRDNDHSDHSRELS
KILAKLERLQONEDLRMAESLRIPREPIDGPGSRAVREDFMKKEIDENKK
QKNGKNDHLEIRRIENGAKELWFLESEKLEKLNENGELORADFLELGHNE
RSIMTDLXYLSOTDAGDWREKAKDLTELVRORYLONPDCKSAKKLYCNINCK
GYCOLHHVYCFMAYAGSPRLAESHMRYATGMPVPRVPSCTDDBSGSSGH
WSGEVKDKVQVVELEPIVDSVPRPRYLALVPEDLADLVRYHGDPAVMYSQPKY
LIRPOWLEKEIEEATKLGFKHPVIGVHVRTRDKVGAFAHPIEEVYHVEDFOL
LARRQVDEKRRVYLAIDPDLAKETKPSYEISDNISSAGLHNRYTENSLENGY
IIDIHFLEQADFLVCTFSSQCVRAVEIEMQALHPDASANFRSLDIYYEGFPNANOI
AIYPHOPRTGEIEIPHEPDIIGVAGNHMDGPKGVNRKIGRTGLVSYVVRKIEIVK
VPTYPEADK"

BASE COUNT 560 a 381 c 443 g 452 t
ORIGIN

Query Match 99.9%; Score 1726.4; DB 4; Length 1836;
Best local Similarity 99.9%; Pred. No. 0;
Matches 1727; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATCGGCGCATGACACTGGTCTGCTGGCGTTGATTCCTCTTTTGGCTGGGGAGCC 60
|||||
Db 57 ATCGGCGCATGACACTGGTCTGCTGGCGTTGATTCCTCTTTTGGCTGGGGAGCC 116
|||||
QY 61 TTGCTATTTTACATAGTGTCTGCTGATGATTAATACACACTGTGACTCTAGC 120
|||||
Db 117 TTGCTATTTTACATAGTGTCTGCTGATGATTAATACACACTGTGACTCTAGC 176
|||||
QY 121 CGAGAACTGTCGAAGATTTTGGCAAGCTGGAAGCTTAACAAACAAATGAAGACTTG 180
|||||
Db 177 CGAGAACTGTCGAAGATTTTGGCAAGCTGGAAGCTTAACAAACAAATGAAGACTTG 236
|||||
QY 181 AGGAGAAAGGCTGATCTCTCCGAATACGGAAGGCCCATTTGATCAGGGGCGAGCTTCA 240
|||||

|||||
Db 237 AGGAGAAAGGCTGATCTCTCCGAATACGGAAGGCCCATTTGATCAGGGGCGAGCTTCA 296
|||||
QY 241 GGAAGACTTGCCTCTTGAAGAGCAATTTATGAAGGCCCAAGACAGATTTGAATTTAT 300
|||||
Db 297 GGAAGACTTGCCTCTTGAAGAGCAATTTATGAAGGCCCAAGACAGATTTGAATTTAT 356
|||||
QY 301 AAGAACAACACTAAATATGCTCCAGGAAAGATCATGAATCTCTAAGAGAGAGATTGAA 360
|||||
Db 357 AAGAACAACACTAAATATGCTCCAGGAAAGATCATGAATCTCTAAGAGAGAGATTGAA 416
|||||
QY 361 AATGAGACTTAAAGAGCTCTGTTTTTCTACAAAGTGAAGATTGAAGAATTTTA 420
|||||
Db 417 AATGAGACTTAAAGAGCTCTGTTTTTCTACAAAGTGAAGATTGAAGAATTTTA 476
|||||
QY 421 GAAGAAATGACATCCAAACACATGCAGATGATTTCTATCAATTTGGGACATCATGAA 480
|||||
Db 477 GAAGAAATGACATCCAAACACATGCAGATGATTTCTATCAATTTGGGACATCATGAA 536
|||||
QY 481 AGGCTATATATGACAGATCTATCTACTACCTCAGTCAACACAGATGGGCGAGTTCGCGT 540
|||||
Db 537 AGGCTATATATGACAGATCTATCTACTACCTCAGTCAACACAGATGGGCGAGTTCGCGT 596
|||||
QY 541 GAAAGAGGCGCAAGATCTGACAGAGCTGTCACGCGGAGAAATTAACATATCTTCAGAT 600
|||||
Db 597 GAAAGAGGCGCAAGATCTGACAGAGCTGTCACGCGGAGAAATTAACATATCTTCAGAT 656
|||||
QY 601 CCAAGAGCTGACAGCAAGGCAAGGCTGTGTGTAATTCAAACAAGCTGTGGCTAT 660
|||||
Db 657 CCAAGAGCTGACAGCAAGGCAAGGCTGTGTGTAATTCAAACAAGCTGTGGCTAT 716
|||||
QY 661 GGCTGTACGCTCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
|||||
Db 717 GGCTGTACGCTCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 776
|||||
QY 721 CTCGCTTGGATCTCACAATTTGGCGCTACGCTACTGGGGGATGGGAACTGTGTTAGA 780
|||||
Db 777 CTCGCTTGGATCTCACAATTTGGCGCTACGCTACTGGGGGATGGGAACTGTGTTAGA 836
|||||
QY 781 CCGTAAAGTGAAGCGTGCACAGACAGATCTGGGAGCTCCACTGAGACATTTGGTACGGTGA 840
|||||
Db 837 CCGTAAAGTGAAGCGTGCACAGACAGATCTGGGAGCTCCACTGAGACATTTGGTACGGTGA 896
|||||
QY 841 GTAAGGACAAATATGTTAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCATCTCGT 900
|||||
Db 897 GTAAGGACAAATATGTTAGTGTGTTGAGCTCCCATTTGTAGACAGTGTTCATCTCGT 956
|||||
QY 901 CCTTCATATTTACCCCTGGCTGTCCAGAAAGACCTTGCAGATGACTTGTAGCAGTTCAT 960
|||||
Db 957 CCTTCATATTTACCCCTGGCTGTCCAGAAAGACCTTGCAGATGACTTGTAGCAGTTCAT 1016
|||||
QY 961 GGGATGCTCAGTGGTGGTATTCCTGATTTGTCAAGTACTTGAATTCGCCCAACCC 1020
|||||
Db 1017 GGGATGCTCAGTGGTGGTATTCCTGATTTGTCAAGTACTTGAATTCGCCCAACCC 1076
|||||
QY 1021 TGGCTGAAAGAAATAGAAAGAGGCCACCAAGAGCTTAGCTTCAAAATTCAGTTATT 1080
|||||
Db 1077 TGGCTGAAAGAAATAGAAAGAGGCCACCAAGAGCTTAGCTTCAAAATTCAGTTATT 1136
|||||
QY 1081 GGAATCATTTAGACGACACAGACAAAGTGGAGCGGAACACGCTTCCATTCATTTGAG 1140
|||||
Db 1137 GGAATCATTTAGACGACACAGACAAAGTGGAGCGGAACACGCTTCCATTCATTTGAG 1196
|||||
QY 1141 GAATPACAGCTGACGTTGAAGAAAGCTTTCAGCTTCTTCTGCGAAGATGCAAGTGCAT 1200
|||||
Db 1197 GAATPACAGCTGACGTTGAAGAAAGCTTTCAGCTTCTTCTGCGAAGATGCAAGTGCAT 1256
|||||
QY 1201 AAAAAAGGCTGATTTTGGCAGACAGATGACCTGCTTTGTTAAAGAGGCAAAAGCAAG 1260
|||||
Db 1257 AAAAAAGGCTGATTTTGGCAGACAGATGACCTGCTTTGTTAAAGAGGCAAAAGCAAG 1316
|||||
QY 1261 TACCCAGTTATGAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
|||||

Db 1317 TACCCAGTTATGAATTTATAGTATTAACCTATCTTGGTCAAGCTGACATATAAT 1376
QY 1321 GATATACGAAATTTACTTGGGGTGTGATCTGGATATACATTTCTCCCGAGCA 1380
Db 1377 GATATACGAAATTTACTTGGGGTGTGATCTGGATATACATTTCTCCCGAGCA 1436
QY 1381 GACTTCTAGTGTGATCTTTTATCGACAGTCTGTAGAGTTGCTTATGAATCATGCA 1440
Db 1437 GACTTCTAGTGTGATCTTTTATCGACAGTCTGTAGAGTTGCTTATGAATCATGCA 1496
QY 1441 GCGTGTGATCTGTATGCTCTCCGAATCTTCCCTTTGGATGACATCTATTTTGA 1500
Db 1497 GCGTGTGATCTGTATGCTCTCCGAATCTTCCCTTTGGATGACATCTATTTTGA 1556
QY 1501 GGGCCAAATGGCCACCAACCAATGTCATTTATCCGACCAACCTGCAAGTGAAGAGA 1560
Db 1557 GGGCCAAATGGCCACCAACCAATGTCATTTATCCGACCAACCTGCAAGTGAAGAGA 1616
QY 1561 ATCCCATGGAACCTGAGATATTTATGCTGTGCTGCAAAATCAGTGGATGCTTCT 1620
Db 1617 ATCCCATGGAACCTGAGATATTTATGCTGTGCTGCAAAATCAGTGGATGCTTCT 1676
QY 1621 AAGGTGTTAACAGAAACTGGGAAGAGCGGCTATATCCCTCTTACAAAGTTGAGAG 1680
Db 1677 AAGGTGTTAACAGAAACTGGGAAGAGCGGCTATATCCCTCTTACAAAGTTGAGAG 1736
QY 1681 AAGATGAAGAACGTCAGATACCCCATATCCCGAGGCTGACAAAGTAA 1728
Db 1737 AAGATGAAGAACGTCAGATACCCCATATCCCGAGGCTGACAAAGTAA 1784

RESULT 4
AF247186 1728 bp mRNA linear MAM 24-APR-2000
LOCUS Bos taurus 6-alpha-L-fucosyltransferase (FUT8) mRNA, complete cds.
DEFINITION AF247186
ACCESSION AF247186.1 GI:7638406
VERSION
KEYWORDS
SOURCE
ORGANISM Bos taurus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 1728)
AUTHORS Javaud,C., Dupuy,F., Maftah,A., Michalski,J.C., Otiol,R.,
Peltic,J.M. and Julien,R.

TITLE Ancestral exon organization of FUT8, the gene encoding the
alpha6-fucosyltransferase reveals successive peptide domains which
suggest a particular 3D core-structure for the
alpha6-fucosyltransferase family

JOURNAL unpublished

REFERENCE 2 (bases 1 to 1728)
AUTHORS Javaud,C., Julien,R. and Peltic,J.M.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2000) Unite de Genetique Moleculaire Animale,
UR1061, Faculte des Sciences, 123, Avenue Albert Thomas, Limoges
87060, France

FEATURES
source Location/Qualifiers
1..1728

gene /organism="Bos taurus"
/db_xref="taxon:9913"
/tissue_type="lung"
1..1728
/gene="FUT8"
1..1728
/gene="FUT8"
/codon_start=1
/product="6-alpha-L-fucosyltransferase"
/protein_id="AF65460.1"
/db_xref="gi:7638407"
/translation="MRPTGSMRWIMLILFAMGTLFLFYIGSHLVBDNDHPHSHSHEL
KILAKLERIKQONEDLRMAESLRIPEDIPDGPASGRIRALEDLVAKEDIEYKK
OTRNLGKIHIEILRRRIENGAKELWFLFQSELKILKLNENGLDHADEFLSDIGHNE

RSIMDLVYLSOTDAGDMREKAEKDLTELVRRTTYLQNPDKDSKAKILVCNINKG
GYGCOLHHVVVCPMAYGORTLILESNNRYATGMEVTRPVSFCTDSDGYTGH
MSGETKDNNOVVELPIYDSLHPRPPIYPLAVPELDRLRVHBDPAVMVYSOPVKY
LIRPOWLEKEIEEATKLGFRHPVIGVHRTBDVGEAFHPLEEVVVEEHPOL
LARRQVOKKRVYLATDPSLKEATKYVPIYEFISDSIMSAGLHRYTENSILGY
ILDIHLEQADFLVCTFSSOVCRAVAYEIMQILHPASNFSLDIDYFPGGNHNOI
AIVHEPRTADEIPEMPEDDIIGVAGNHMDGSKVNRKLGRTGLPSYKVEKIEYK
VPHPEAK"

BASE COUNT 517 a 368 c 418 g 425 t
ORIGIN
Query Match 89.4%; Score 1545.6; DB 4; Length 1728;
Best Local Similarity 93.4%; Pred. No. 0;
Matches 1614; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 1 ATGCGGCATGAGACTGCTGCTGGCGCTTGATTAAGCTCATCTTTTGGCGGGACC 60
Db 1 ATGCGGCATGAGACTGCTGCTGGCGCTTGATTAAGCTCATCTTTTGGCGGGACC 60
QY 61 TTGCTATTTTACATAGCTGTGCTGCTGTGAGAGATTAAGACACTGTGACTAGC 120
Db 61 TTGCTATTTTACATAGCTGTGCTGCTGTGAGAGATTAAGACACTGTGACTAGC 120
QY 121 CGAGAACTGTCCAAAGATTTTGGCAAGCTGAAACCTTAAACCAACAAATGAAGACTTG 180
Db 121 CGAGAACTGTCCAAAGATTTTGGCAAGCTTGAACCTTAAACCAACAAATGAAGACTTG 180
QY 181 AGGAGAACTGTGAGATCTCTCCGAATACGAGAGGCCCATTTGATCAGGCGCAGCTTCA 240
Db 181 AGGAGAACTGTGAGATCTCTCCGAATACGAGAGGCCCATTTGATCAGGCGCAGCTTCA 240
QY 241 GGAAGAGTCTGCTTAAAGAGCAATTTATGAAGGCCCAAGAACAGATTGAATTTAT 300
Db 241 GGAAGAGTCTGCTTAAAGAGCAATTTATGAAGGCCCAAGAACAGATTGAATTTAT 300
QY 301 AAGAAACAACATAAAATGCTCCAGGAGAGATCATGAATCTTAAGAGAGAGATTGAA 360
Db 301 AAGAAACAACATAAAATGCTTGGGAGAGATCATGAATCTTAAGAGAGAGATTGAA 360
QY 361 AATGAGCTAAAGAGCTGTGTTTTTCTTACAAAGTGAAGTGAAGAAATTAAGAATTTA 420
Db 361 AATGAGCTAAAGAGCTGTGTTTTTCTTACAAAGTGAAGTGAAGAAATTAAGAATTTA 420
QY 421 GAAGAAATGAAGCTCCAAAGACATGAGATGAATTTCTATGAGATTGGACATCTGAA 480
Db 421 GAAGAAATGAAGCTCCAAAGACATGAGATGAATTTCTATGAGATTGGACATCTGAA 480
QY 481 AGCTATATATGACGATCTATATCTACGTCAGTCAACAGATGAGACAGTGTGGCGG 540
Db 481 AGCTATATATGACGATCTATATCTACGTCAGTCAACAGATGAGACAGTGTGGCGG 540
QY 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCCAGCGGAGAAATTAATATCTTCAAT 600
Db 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCCAGCGGAGAAATTAATATCTTCAAT 600
QY 601 CCCAAGAGATGAGCAACCAAGAGCTGTGTAATATCAACCAAGAGCTGTGCTAT 660
Db 601 CCCAAGAGATGAGCAACCAAGAGCTGTGTAATATCAACCAAGAGCTGTGCTAT 660
QY 661 GGCTGTACCTCATCATCTATGATGATCTGCTTATGATGATGATGATGATGATGATGAT 720
Db 661 GGCTGTACCTCATCATCTATGATGATCTGCTTATGATGATGATGATGATGATGATGAT 720
QY 721 CTGCTGTGGAATCTCAATTTGGCGCTAGCTAGTGGGAGATGGGAACTGTGTTTGA 780
Db 721 CTGCTGTGGAATCTCAATTTGGCGCTAGCTAGTGGGAGATGGGAACTGTGTTTGA 780
QY 781 CCTGTAACTGAGACGTGACACAGATGTGCGACGTCTCACTGACATTTGTCAGGTGA 840
Db 781 CCTGTAACTGAGACGTGACACAGATGTGCGCTTACACTGACACTGTCAGGTGA 840
QY 841 GTAAAGAGCAAAATGTTAGGTGAGTCCCATGTTGAGACAGTGTGATGCTGCTGT 900
Db 841 GTAAAGAGCAAAATGTTAGGTGAGTCCCATGTTGAGACAGTGTGATGCTGCTGT 900

Db 841 ATAAAGACAAAAATGTACAGTAGTCGAGCTCCCATTTGTGGACAGTCTTACCTCGT 900
QY CCTCCATATTTTACCCTGGCTGTCCAGAGACCTTGCAGATGACGTGTAGAGTCAT 960
Db 901 CCGTCATATCTACCCCTGGCTGTACCAAGAACCTTGCAGATGGCTTGTACAGTCCAT 960
QY 961 GGTGATCTGTCAGATGTGGGGATATCCAGTTTGTCAAGTACTTGTATTCGCCACAACC 1020
Db 961 GGTGATCTGTCAGATGTGGGGATATCCAGTTTGTCAAGTACTTGTATTCGCCACAACC 1020
QY 1021 TGGCTGGAAGAAATAGAGAGGCCACAGAGCTAGCTTCAACATCCAGTTAT 1080
Db 1021 TGGCTGGAAGAAATAGAGAGGCCACAGAGCTAGCTTCAACATCCAGTTAT 1080
QY 1081 GGAGTCCATGTTAGACGACACAGCAAGTGGGCGGAAGCAGCTTCCATTCATTTAG 1140
Db 1081 GGAGTCCATGTTAGACGACACAGCAAGTGGGCGGAAGCAGCTTCCATTCATTTAG 1140
QY 1141 GAATACACGCTGACGCTTGAAGAAGACTTTCAGCTTCTTGGCTGCGAGAAATGCAAGTGAT 1200
Db 1141 GAATACACGCTGACGCTTGAAGAAGACTTTCAGCTTCTTGGCTGCGAGAAATGCAAGTGAT 1200
QY 1201 AAAAAAGGTGATTTGGCCACAGATGACCTTCTTTGTTAAAGAGCAAAACAAAG 1260
Db 1201 AAAAAAGGTGATTTGGCCACAGATGACCTTCTTTGTTAAAGAGCAAAACAAAG 1260
QY 1261 TACCCCATTTATTAATTTATTTAGTAACTATCTATCTTGGTGCAGCTGACATCAAT 1320
Db 1261 TACCCCATTTATTAATTTATTTAGTAACTATCTATCTTGGTGCAGCTGACATCAAT 1320
QY 1321 CGATATACAGAAATTCACCTTGGGGTGTGATCCTGATATACACTTCTCTCCAGCA 1380
Db 1321 CGATATACAGAAATTCACCTTGGGGTGTGATCCTGATATACACTTCTCTCCAGCA 1380
QY 1381 GACTTCTAGTGTACTTCTTTCATGCGAGCTGTAGAGTCTTATGTAATCAATGCA 1440
Db 1381 GACTTCTAGTGTACTTCTTTCATGCGAGCTGTAGAGTCTTATGTAATCAATGCA 1440
QY 1441 GCGCTCATCTGATGCTCTGCGAATCTCGTCTTGGTGAACATCTCTATTTTGA 1500
Db 1441 GCGCTCATCTGATGCTCTGCGAATCTCGTCTTGGTGAACATCTCTATTTTGA 1500
QY 1501 GGGCCAAATGCCACAAATTTGCCATTATCTTCACCACTCGAATGAGAGAA 1560
Db 1501 GGGCCAAATGCCACAAATTTGCCATTATCTTCACCACTCGAATGAGAGAA 1560
QY 1561 ATCCCATGGAACCTGAGATATTTGGTGGCTGGAATACAGTGGATGCTATCT 1620
Db 1561 ATCCCATGGAACCTGAGATATTTGGTGGCTGGAATACAGTGGATGCTATCT 1620
QY 1621 AAGGTGTTAAAGAAAGTGGGAGGAGGCTATATCTCTCAACAAGTTGAGAG 1680
Db 1621 AAGGTGTTAAAGAAAGTGGGAGGAGGCTATATCTCTCAACAAGTTGAGAG 1680
QY 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGCTGACAGATTA 1728
Db 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGCTGACAGATTA 1728

RESULT 5
Db9289
LOCUS Db9289 2002 bp mRNA linear PRI 07-FEB-1999
DEFINITION Homo sapiens mRNA for N-Acetyl-beta-D-glucosaminide, complete cds.
ACCESSION D89289.1 GI:2055306
VERSION 1
KEYWORDS N-Acetyl-beta-D-glucosaminide; GDP-L-Fuc; alpha 1-6
SOURCE Fucosyltransferase: alpha1-6 FUCT.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
Yanagidani,S., Uozumi,N., Ihara,Y., Miyoshi,E., Yamaguchi,N. and

Taniguchi,N.
Purification and cDNA cloning of
GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alpha1-6 fucosyltransferase
(alpha1-6 Fuct) from human gastric cancer MKN45 cells
J. Biochem. 121 (3), 626-632 (1997)
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
CDs
FEATURES
1.2002
/organism="Homo sapiens"
/db.xref="taxon:9606"
143.1870
/note="GDP-L-Fuc; alpha 1-6 Fucosyltransferase; alpha1-6
Fuct"
/product="N-Acetyl-beta-D-glucosaminide"
/protein_id="BA19764.1"
/db.xref="GI:2055307"
/translation="MRPWGSMRWIMLILFWAGTLEFYIGHLVNDNDIPDHSRELS
KILAKLERKQONEDLRMAESLRIPEDIDGPAIGRVLEEDLVKAEQIDENYK
OTRNGKXKHEILRRRIENGAKELMFLOSELKLNLEGNELORHADFLLDNE
RSIMDLVYLSOTDAGAREKELADLELVORITLYLONPKDSKAKKLVGNINIKG
GYGCOLHNHYVCFMAYTCQRTLILLESOMWRATGWEIVFPVSECTDRSGISGCH
WSGEVKNVQVVELPVDLSLRPRPYLPLAVBEDLVLVHEDPAVWMSQVRY
LIRPQWLEKELEATKLGFKHPIVGVHRTDVGTEAAHPIEEVWVHEEHOL
LARRQVDEKRYVLATDDPSLKEAKTKYRNEYISDSISWAGHINRYTENSJRGV
LIDHIFLSQADFLVCTFSOVCRAVEIMOTLHPDASNFHSLDILVYFGONAHNOI
ALVADHPRTRADLEIPMEPDILGVAGHNMDSKGVNRKLGRTGLVPSYKVRKELFTVK
YFTYPEAEK"

BASE COUNT 620 a 415 c 467 g 500 t
ORIGIN
Query Match 87.4%; Score 1510.4; DB 9; Length 2002;
Best local Similarity 92.1%; Pred. No. 0;
Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 1 ATCGGGCCATGAGCTGGTGGCTGGGCTTGAATATGCTATTCCTTTTGGCTGGGAGC 60
Db 143 ATCGGGCCATGAGCTGGTGGCTGGGCTTGAATATGCTATTCCTTTTGGCTGGGAGC 202
QY 61 TTGCTATTTTACATAGTGTGCTCACTTGTGTACGAGATTAATACCACTGTACTAGC 120
Db 203 TTGCTATTTTATAGTGTGCTCACTTGTGTACGAGATTAATACCACTGTACTAGC 262
QY 121 CGAAGCTGTCCAAAGATTTTGGCAAAAGCTGGAACGCTTAACAACAATAATGAAGACTTG 180
Db 263 CGAAGCTGTCCAAAGATTTTGGCAAAAGCTGGAACGCTTAACAACAAGATTAAGACTTG 322
QY 181 AGGAGATGCTGATCTCTCCGATACAGAAAGGCCCATTTGATCAGGGGCGAGCTTCA 240
Db 323 AGGAGATGCTGATCTCTCCGATACAGAAAGGCCCATTTGATCAGGGGCGAGCTTCA 382
QY 241 GGAAGATGCTGCTTTAGAGAGCAATTTATGAAGGCCAAAGACAGATTTGAAATTTAT 300
Db 383 GGAAGATGCTGCTTTAGAGAGCAATTTATGAAGGCCAAAGACAGATTTGAAATTTAT 442
QY 301 AAGAAACAACTAAATAGTTCAGGAGAGATCATGAATCTTAAGAGAGAGATTTGAA 360
Db 443 AAGAAACAACTAAATAGTTCAGGAGAGATCATGAATCTTAAGAGAGAGATTTGAA 502

Db	1563	ACACTACATTCCTGATGCTCTGTGCAAACTTCATTCCTTTAGATGACATCTACTATTTTGGG	1642
QY	1501	GGCCCAATGCCCACCAACCAAAATGGCCATTTATGCTCCACCAACCTCGAACTGAAGAGAA	1560
Db	1643	GGCCAGAAATGCCCACCAATCAAAATTTAGCCATTTATGCTCCACCAACCCCGAACTCGAGATGAA	1702
QY	1561	ATCCCCATGGAACCTGGAGATATTATTGGTGTGGCTGGAATTCACCTGGATGCTATTCCT	1620
Db	1703	ATTCCTCAATGGAACCTGGAGATATTATTGGTGTGGCTGGAATTCATTGGGATGGCTATTCT	1762
QY	1621	AAAGGTGTATACAGAAAACCTGGGAAGGACGGGGCCATATTCGCTCCTCCATCAAGTTCGAGAG	1680
Db	1763	AAAGGTGTCAACAGAAATTTGGGAAGGACGGGGCCATATTCCTCCTCCATCAAGATTCGAGAG	1822
QY	1681	AAAGTACGAACAGTCGAATACCCACATATTCGACGAGTACGAAGTAA	1728
Db	1823	AAAGTACGAACAGTCGAATACCCACATATTCGAGCTGAGGAATTA	1870
RESULT 6			
LOCUS	E15725	2100 bp	DNA
DEFINITION	Human mRNA for alpha-1,6-fucosyltransferase, complete cds.		linear PAT 28-JUL-1999
ACCESSION	E15725		
VERSION	E15725.1	GI:5710408	
KEYWORDS	JP 1998084975-A/1.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2100)		
TITLE	Taniguchi, N., Uozumi, H., and Yanagiya, S.		
JOURNAL	ALPHA-1,6-FUCOSYLTRANSFERASE GENE DERIVED FROM HUMAN Patent: JP 1998084975-A 1 07-APR-1998; TOYOBO CO LTD		
COMMENT	OS Homo sapiens (human) PN JP 1998084975-A/1 PD 07-APR-1998 PE 17-JUN-1997 JP 1997159692 PR 22-JUL-1996 JP 96P 199260 PI TANIGUCHI NAOUKI, UOZUMI HISAUMI, YANAGIYA SHUSAKU PC C12N15/09, C07H21/04, C12N1/21, C12N9/10, C12N15/09, C12H1/91, PC C12N1/21, PC C12R1/19, C12N9/10, C12R1/19); CC strandedness: Double; CC topology: Linear; FH key Location/Qualifiers FH source 1. 2100 FT /organism='Homo sapiens' FT /cell_type='stomach cancer cell' FT FT /cell_line='MKM45', FT CDS 198..1925 FT /product='alpha-1,6-fucosyltransferase'; FEATURES source 1. 2100 Location/Qualifiers 1. 2100 /organism='Homo sapiens' /db_xref='taxon:9606' BASE COUNT 651 a 436 c 489 g 524 t ORIGIN Query Match 87.4%; Score 1510.4; DB 6; Length 2100; Best Local Similarity 92.1%; Pred. No. 0; Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;		
QY	1	ATGCGGCCATGAGCTGTTGCTGGCGGTGGATATGCTCATTTCTTTTGGCTGGGGGACG	60
Db	198	ATGCGGCCATGAGCTGTTGCTGGCGGTGGATATGCTCATTTCTTTTGGCTGGGGGACG	257
QY	61	TTGCTATTTTACATAGGTGCTGCTCTGGTACGAGATATGACCACTCTGATCAGCTTACG	120
Db	258	TTGCTATTTTATATAGGTGCTGCTCTGGTACGAGATATGACCACTCTGATCAGCTTACG	317

Qy	121	CGAAGACTGTCGAAGATTTTGGCCAAAGCTGGAAGCCTTTAAACCAACAAATGSAAGACTTG	180
Dp	318	CGAAGACTGTCGAAGATTTTGGCCAAAGCTTGAAGCGTTTAAACACACAGATGAAGAACTTG	377
Qy	181	AGGAAATGCGCTGGATCTCTCCGAATACGAAGAGCCCATTTGATCAGGGGCACTTC	240
Dp	378	AGGGAAATGCGCCAAATCTCTCCGGATACCGAAGAGGCCATTTGATCAGGGGCACTTC	437
Qy	241	GGAGAGCTTGGTCTTTTATGAGAGCAATTTATGAMGGCCAAAGAACAGATTGAATTAAT	300
Dp	438	GGAGAGAGTACCGCTTTTATGAGAGAGCAGCTTTGTTAAAGCCAAAGAACAGATTGAATTAAT	497
Qy	301	AGAAGAACAACTAAAAATTTGGTCCAGGAAAGATCATGAATCTTAAGGAGGAGTTTGA	360
Dp	498	AGAAGAACAGACACGAATTTGGTCTGGGGAAGAGATCATGAATCTTGAAGAGAGGATTGA	557
Qy	361	AATGAGACTAAAGAGCTCTGGTTTTTTTCTACAAAGTGAAGTGAAGAAATTAAGAAATTA	420
Dp	618	GAAGAAATGAACCTCCAAACACATGCAAGATTTCTATCAGATTTGGGACATCATGA	480
Qy	481	AGGCTATTAATGACGATCTATCTACTACCTCAGTCAAAACAGATGGGGCAGGTGATGGCGT	540
Dp	678	AGGCTATTAATGACGATCTATCTACTACCTCAGTCAAAACAGATGGGATGATGGCGG	737
Qy	541	GAAAGAGAGGCCAAAGATCTGACAGAGCTGCTCCAGCGAGAAATTAACATATCTTGACAT	600
Dp	738	GAAAGAGAGGCCAAAGATCTGACAGAGAGCTGCTCCAGCGAGAAATTAACATATCTTGACAT	797
Qy	601	CCCAAGGACTGACACAAAGCCAAAGAGCTAAGTGTGAATATCAACAAAGGCTGTGGCTAT	660
Dp	798	CCCAAGGACTGACACAAAGCCAAAGAGCTAAGTGTGAATATCAACAAAGGCTGTGGCTAT	857
Qy	661	GGCTGTGAGCTCCATCATGTAGTGTACTGCTCTTATGATTTGCATATGACACCCAGGAACA	720
Dp	858	GGCTGTGAGCTCCATCATGTAGTGTACTGCTCTTATGATTTGCATATGACACCCAGGAACA	917
Qy	721	CTCGCCTTTGGAATCTGCATTTGGGCGCTACGCTACAGGGGGATGGGAAACGTGTTTGA	780
Dp	918	CTCATCTTGGGAATCTGCATTTGGGCGCTATGCTACTGAGTGGAGTGGAGACTGTATTTAGG	977
Qy	781	CCCTGAAGTGAAGAGCTGACAGACAGATCTGGCAGCTCCACTGCGACATTTGGTCAAGTGA	840
Dp	978	CCCTGAAGTGAAGAGCTGACAGACAGATCTGGCAGCTCCACTGCGACACTGGTCAAGTGA	1037
Qy	841	GTTAAGGACAAAATGTTCAAGTGTGATGAGCTCCCATTTGTAGACAGTGTATCTCTGCT	900
Dp	1038	GTTAAGGACAAAATGTTCAAGTGTGATGAGCTCCCATTTGTAGACAGTGTATCTCTCTGCT	1097
Qy	901	CCCTCATATTTTACCCCTGGGCTGTCCCAAGAAAGCTTGCAGATCTGTACAGATTCAT	960
Dp	1098	CCCTCATATTTTACCCCTGGGCTGTACCAAGAAAGCTTGCAGATCTGTACAGATTCAT	1157
Qy	961	GGTGAATCCCTGCACTGTGTGGGTGATCCCAAGTGTGCAAGTACTGATTTGGCCCAACACC	1020
Dp	1158	GGTGAATCCCTGCACTGTGTGGGTGATCCCAAGTGTGCAAGTACTGATTTGGCCCAACACC	1217
Qy	1021	TGGGTGAAAAAGAAATATGAAGAGGCGCACCAAGAGCTGTAGGCTTCAAAACATCCAGTTAT	1080
Dp	1218	TGGGTGAAAAAGAAATATGAAGAGGCGCACCAAGAGAGCTGTAGGCTTCAAAACATCCAGTTAT	1277
Qy	1081	GGAGTCCCATGTTTAAAGCAGACAGACAAAGTGGGAGCGGAAGCAGCCTTCCATCCATTGAG	1140
Dp	1278	GGAGTCCCATGTTTAAAGCAGACAGACAAAGTGGGAGCAAGAAAGTGTCCATCCATTGAA	1337
Qy	1141	GAATATACAGGTGCACGTTTGAAGAAAGCTTTAGCTTTCTTGCTGCGAGAAATGCAAGTGCAT	1200
Dp	1338	GAGTATATGTTGCACTGTTTGAAGAAAGCTTTAGCTTTCTTGCTGCGAGAAATGCAAGTGCAT	1397

OY	1201	AAAAAAGGGTGTATTGTTGGCCAGCAATATGACCTCGTTTGTTTAAAGAGGCCAAAAACAAG	1260
OY	1201	AAAAAAAAAGGGTGTATTGTTGGCCAGCAATATGACCTCGTTTGTTTAAAGAGGCCAAAAACAAG	1260
Db	1398	AAAAAAGAGTGTATTGTTGGCCAGCAATATGACCTCGTTTGTTTAAAGAGGCCAAAAACAAG	1457
OY	1261	TACCCAGTTATGATTTATTTAGTATTAACCTATCTCTTGCTGAGCTGACATACATAT	1320
Db	1458	TACCCCAATTATGAATTTTATTTAGTATTAACCTATCTCTTGCTGAGCTGACATACATAT	1517
OY	1321	CGATATACAGAAAATTCACCTTGGGGGTGTGATCCTGATATATACACTTTCCTGCCAGCA	1380
Db	1518	CGATACACAGAAAATTCACCTTGGGGGTGTGATCCTGATATATACACTTTCCTGCCAGCA	1577
OY	1381	CACATTCCTAGTGTGTGACTTTTTCATCCGACAGCTCTAGAGTTGCTTATGAATTCATTCGA	1440
Db	1578	GACTTCCTAGTGTGTGACTTTTTCATCCGACAGCTCTAGAGTTGCTTATGAATTCATTCGA	1637
OY	1441	GGCGTCATCTCGATGAGCCCTGCGAAGCTTCGCTTCTTGATGACATCGATCATTTTGGGA	1500
Db	1658	ACACTACATCTCGATGAGCCCTGCGAAGCTTCGCTTCTTGATGACATCGATCATTTTGGGA	1697
OY	1501	GGCCCAATGCCCCACACCAAAATTCGCAATTTATCTCACCACCACTCGAAGCAAGAGAA	1560
Db	1698	GGCCGAGAAATGCCCAACATCAATATTCGCAATTTATCTCACCACCACTCGAAGCAAGAGAA	1757
OY	1561	ATCCCATGGAACCTCGAGATATTTTGTGTGCTGCGAATCACTGGGATGGCTATCTCT	1620
Db	1758	ATTCCTCATGGAACCTCGAGATATTTTGTGTGCTGCGAATCACTGGGATGGCTATCTCT	1817
OY	1621	AAAGCTGTTAAACAGAAAACCTGGGAGGAGGAGGCTATATCTCTCTACAAAGTTGCGAG	1680
Db	1818	AAAGCTGTTAAACAGAAAACCTGGGAGGAGGAGGCTATATCTCTCTACAAAGTTGCGAG	1877
OY	1681	AAAGTGAAGAACAGTCAAGTACCCCATATATCCGAGGCTGACAACTAA	1728
Db	1878	AAAGTGAAGAACAGTCAAGTACCCCATATATCTGAGGCTGACAACTAA	1925
RESULT 7	HSY17976	2898 bp	linear
LOCUS	HSY17976	2898 bp	linear
DEFINITION	Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase, transcript B1.	2898 bp	linear
ACCESSION	Y17976	GI:3451262	
VERSION	Y17976.1	GI:3451262	
KEYWORDS	FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 2898) Cailliau, A., Balazs, L., Candelier, J. J., Oriol, R. and Mollicone, R.		
TITLE	Differential splice variants of human FUT8 embryonic cDNA		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2898)		
AUTHORS	Cailliau, A. L. V.		
TITLE	Direct Submission		
JOURNAL	Submitted (19-AUG-1998) A. L. V. Cailliau, INSERM U178, 16 av P. V. Courier, 94807 Villejuif Cedex, FRANCE		
COMMENT	Related entry: DB9289.		
FEATURES	Location/Qualifiers		
SOURCE	1. .2898		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="14"		
	/map="14q23"		
	/clone="FUT8"		
	/tissue="Whole embryo"		
	/dev_stage="40-70 days"		
	1. .2874		
	/gene="FUT8"		
	1. .101		
	/gene="FUT8"		
gene	exon		

```
/note="B"
102..200
/gene="FUT8"
/note="C"
201..2156
/gene="FUT8"
/note="D"
429..2156
/gene="FUT8"
/EC_number="2.4.1.68"
/note="allele 1"
/codon_start=1
/product="glycoprotein 6-alpha-L-fucosyltransferase"
/protein_id="CAA76985.1"
/db_xref="GI:3451263"
/db_xref="SPTREMBL:O00235"
/translation="MRPTGSMRWITMLIFAWGTLLEFYIGHLVRNDHPDHSHNELS
KILAKLERIKOONEDLRMAESLRIPEDPIDGPAIGHVRVLEOLVAKEDIEYKK
OTRNLGKDHEILRRRIENGAKELFLOESLKLKLENGELORHADEFLDLNHE
RSIMDLVYLSOTDGDMDREKAOLIELVORRTYVONPKDSKAKKYNINIKGC
GYGCOLHVYVCFMIAVSTORTLILESNNMRYATGMEVTVRPVSECTDMSIGTGH
WSGEVKDNVOVLELPIVDSLHPREPVLPVAVPELDRLVRHVDPAVMVVSQFVKY
LIRQPMLEKEIEEATKRLGFKHPYIGVHRRTRDVGDEAFHPLEEMVHVEEHL
LARMOVDKRRVYLATDDPSLKEAKTYRPNVEFSDNSISWSAGLHNRYESNLGV
LIDHFLSQADFLVCTFSSOYCRVAEIMOTLHPDASANPFLSDIYIEGGGNHNOI
ALYAHOPRTADEIPEPBDIIGVAGNHMDGVSKVNRRLGRTGLPSYKVRKEITVK
VPTPEAEK"
variation
2218
/gene="FUT8"
/note="I1"
/frequency=".5"
/replacement="C"
2758..2763
/gene="FUT8"
/note="putative"
2770..2775
/gene="FUT8"
/note="putative"
2869..2874
/gene="FUT8"
/note="putative"
BASE COUNT      881 a      589 c      641 g      787 t
ORIGIN
Query Match      87.4%; Score 1510.4; DB 9; Length 2898;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
1 ATGCGGCATGAGACTGGTCTGCGGCGTTGGATTATGCTCATTTCTTTCCTGGGGGACC 60
|||||
429 ATGCGGCCATGAGACTGGTCTGCGGCGTTGGATTATGCTCATTTCTTTCCTGGGGGACC 488
|||||
61 TTGCTATTTTACATAGAGTGTGCTGCTGTAGACGATTAATGACCACCTGTATCTAGC 120
|||||
489 TTGCTGTTTTATATAGTGTGCTGCTGTAGACGATTAATGACCACCTGTATCTAGC 548
|||||
121 CGAGAACTGTCCAAGTTTTTGGCAAGCTGGAACGCTTAAACACAAATGAAGCTTG 180
|||||
549 CGAGAACTGTCCAAGTTTTTGGCAAGCTTGAACGCTTAAACACGAGAAATGAAGCTTG 608
|||||
181 AGGAGAACTGTGATCTCTCCGAATACAGAAAGGCCCATTTGATCAGGGCCAGCTTCA 240
|||||
609 AGCGCAATGCGCAATCTCTCCGAGATACAGAAAGGCCCATTTGATCAGGGCCAGCTTCA 668
|||||
241 GGAAGACTTCTGCTTTTGAAGACAAATTTATGAAGGCCCAAGAACAGATTGAAAATTTAT 300
|||||
669 GGAAGACTTCTGCTTTTGAAGACAAATTTATGAAGGCCCAAGAACAGATTGAAAATTTAC 728
|||||
301 AAGAAACAACATAAATAATGTCAGGAAAGATCATGAATCTTAAGAGAGAGATTTGAA 360
|||||
729 AAGAAACAACATAAATAATGTCAGGAAAGATCATGAATCTTAAGAGAGAGATTTGAA 788
|||||
361 AATGAGCTAAAGAGCTCTGTTTTTCTACAAAGTGAAGTGAAGAAATTTAAAGAAATTTA 420
```

```
|||||
789 AATGAGCTAAAGAGCTCTGTTTTTCTACAGAGGAAATGGAAGAAATTTAAAGAACTTTA 848
|||||
421 GAAGAAATGAACCTCCMAAGACATGAGATGAATTTCTATCAGATTTGGGACATCATGAA 480
|||||
849 GAAGAAATGAACCTCCMAAGACATGAGATGAATTTCTTGTGATTTAGACATCATGAA 908
|||||
481 AGGCTATATATGAGGATTTCTATCTACCTGATCAAGACAGATGGGGACAGTATTTGGCGT 540
|||||
909 AGGCTATATATGAGGATTTCTATCTACCTGATCAAGACAGATGGGGACAGTATTTGGCGG 968
|||||
541 GAAAGAGAGCCAAAGATTTGACAGAGCTGGTCCAGCGGAGATTAATCATATCTTCGAAT 600
|||||
969 GAAAGAGAGCCAAAGATTTGACAGAGCTGGTCCAGCGGAGATTAATCATATCTTCGAAT 1028
|||||
601 CCCAAGAGCTGACAGCAAAAGCCAAAGAGCTAGTGTGAATATCAACAAGAGCTGTGCTAT 660
|||||
1029 CCCAAGAGCTGACAGCAAAAGCCAAAGAGCTGTGTGAATATCAACAAGAGCTGTGCTAT 1088
|||||
661 GGCTGTACAGCTCCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
|||||
1089 GGCTGTACAGCTCCATCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1148
|||||
721 CTCGCTTGGAATCTGCAATTTGAGCGCTGACGCTACTGGGGGATGGGAACCTGTGTTTGA 780
|||||
1149 CTCATCTTGGAAATCTGAGAAATTTGGCGCTATGCTACTGATGATGAGAGACTGATATTAG 1208
|||||
781 CCTGTAGTGAAGAGCTGACAGACAGATCTGGCAGCTCCAGCTGAGATTTGGTCAAGTGA 840
|||||
1209 CCTGTAGTGAAGAGCTGACAGACAGATCTGGCAGCTCCAGCTGAGATTTGGTCAAGTGA 1268
|||||
841 GTTAAAGACAAAATTTTGAAGTGTGAGCTGCCATTTGTAGACAGTGTTCATCTCGT 900
|||||
1269 GTTAAAGACAAAATTTTGAAGTGTGAGCTGCCATTTGTAGACAGTGTTCATCTCGCT 1328
|||||
901 CCTCATATTTACCCCTGGCTGCTCCAGAAAGCTTGCAGATCGATTTGACAGATCAT 960
|||||
1329 CCTCATATTTACCCCTGGCTGCTCCAGAAAGCTTGCAGATCGATTTGACAGATCAT 1388
|||||
961 GGTGATCTGACAGTGTGAGTGTATCCAGATTTGTCAAGTACTTGAATTCGCCCAACACC 1020
|||||
1389 GGTGATCTGACAGTGTGAGTGTATCCAGATTTGTCAAGTACTTGAATTCGCCCAACACC 1448
|||||
1021 TGGCTGGAATAAGAAATTAAGAAGGCCACCAAGAACTGAGCTTCAAAATTCAGTTAT 1080
|||||
1449 TGGCTGGAATAAGAAATTAAGAAGGCCACCAAGAACTGAGCTTCAAAATTCAGTTAT 1508
|||||
1081 GGAATCATGTTAGAGCAGACAGCAAAAGTGGGAGCGGAACAGACCTTCCATCCATTTGAG 1140
|||||
1509 GGAATCATGTTAGAGCAGACAGCAAAAGTGGGAGCGGAACAGACCTTCCATCCATTTGAG 1568
|||||
1141 GAATACAGCTGACAGCTTGAAGAAGCTTTCAGCTTCTTCTCGCAGAAATGCAAGTGAAT 1200
|||||
1569 GAATACAGCTGACAGCTTGAAGAAGCTTTCAGCTTCTTCTCGCAGAAATGCAAGTGAAT 1628
|||||
1201 AAAAAAAGGATGATTTGGCCACAGATGACCTGCTTGTAAAGAGCAAAAAACAAAG 1260
|||||
1629 AAAAAAAGGATGATTTGGCCACAGATGACCTGCTTGTAAAGAGCAAAAAACAAAG 1688
|||||
1261 TACCCAGTTTGAATTTTATGATGATTAATCTATCTTGGTGCACCTGAGATCAATTAAT 1320
|||||
1689 TACCCAGTTTGAATTTTATGATGATTAATCTATCTTGGTGCACCTGAGATCAATTAAT 1748
|||||
1321 CGATATACAGAAATTCACCTTGGGGGTGTGATCTGGAATATACATCTTCTCCAGGCA 1380
|||||
1749 CGATATACAGAAATTCACCTTGGGGGTGTGATCTGGAATATACATCTTCTCCAGGCA 1808
|||||
1381 GACTTCCTAGTGTGATCTTTCATGCGAGCTGTGAAGTGTGATGATGATGATGATGATGAT 1440
|||||
1809 GACTTCCTAGTGTGATCTTTCATGCGAGCTGTGAAGTGTGATGATGATGATGATGATGAT 1868
|||||
1441 GCGCTCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
|||||
```

Db 1869 ACACATACATCCGTGATGCTCTGCAAACTTCCATCTTTAGATGACATCTACTATTTTGGG 1928
OY 1501 GGCCCAATGCCCAACCAATTTGCATTTATCTCACCACCTGCAACTGAAGAGAA 1560
Db 1929 GGGCAGATGCCCAATCAATTTGCCATTTATGCTCACCACCCGAACTGCAGATGAA 1988
OY 1561 ATCCCATGGAACCTGAGATATTTATGATGCTGGAATCATCGATGGCTATCTCT 1620
Db 1989 ATTCCCATGGAACCTGAGATATCTGATGCTGGAATCATCTGGATGGCTATCTCT 2048
OY 1621 AAAGGTCTTAACGAAACCTGGAAGAGAGGCGCTATATCCCTCTCAAAAGTTCCAGAG 1680
Db 2049 AAAGGTCTTAACGAAATTTGGAAGAGAGGCGCTATATCCCTCTCAAAAGTTCCAGAG 2108
OY 1681 AAGATGAAACAGTCAAGTACCCACATATCCGAGGCTGCAAGTAA 1728
Db 2109 AAGATGAAACGCTCAAGTACCCACATATCTCTGAGGCTGCAAGTAA 2156
RESULT 8
LOCUS HS17977 2992 bp mRNA linear prt 17-sep-1998
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase,
transcript B2.
ACCESSION Y17977.1 GI:3451264
VERSION Y17977
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 2992)
Callieau,A., Balanzino,L., Candelier,J.J., Oriol,R. and
Mollicone,R.
TITLE Differential splice variants of human FUT8 embryonic cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2992)
Callieau,A.L.V.
AUTHORS Direct Submission
TITLE Submitted (19-AUG-1998) A.L.V. Callieau, INSERM U178, 16 av P.V.
JOURNAL Coururier, 94807 Villejuif Cedex, FRANCE
COMMENT Related entry: D89289.
FEATURES
location/Qualifiers
source 1..2992
/organism="Homo sapiens"
/isolate="transcript B2"
/db_xref="taxon:9606"
/map="14q23"
/chromosome="14"
/issue_type="whole embryo"
/dev_stage="40-70 days"
1..2874
/gene="FUT8"
1..101
/gene="FUT8"
/note="B"
102..200
/gene="FUT8"
/note="C"
201..>2156
/gene="FUT8"
/note="D"
429..2156
/BC_number="2.4.1.68"
/BC_number="2.4.1.68"
/note="allele 1"
/codon_start=1
/product="glycoprotein 6-alpha-L-fucosyltransferase"
/protein_id="CAA76986.1"
/db_xref="GI:3451265"
/db_xref="SPTREMBL:O00235"
/translation="MRPWTGSMRWIMLILFANGTLLFYIGHLVRDNDHPDHSRELS
KILAKLERLKQONEDLRMAESLRIPGPIIDGPAIGRVVLEEDLVKAKEOIENYKK

OTRNLGKDHDEILRRRIENGAKELMFPELSEELKLRNLNEGNELORIHADEFLLDLCNHE
RSIMIDLYLVSQDGAQWREKAKDLDEFLVYORITLYLONPKDCSKARKLYONIKKGC
GYGCOLDHVYVCFMIAVGTQPTLILSOSMRYATGMEVFPVSESTCDREGISYGH
WSGEYKDKNOYVELPIYDSLHPRPYLPLAEPEDLADRLVPHVDDPAMVMSOPVYK
LIRPOYLEKEIEEATRKIGFNHPYIGHVVRTDYGTBAHPLEEVHYHEEPOL
LARRNOVDKRRVLYATDDPSLKEAKTKYPNTEISDNSISAGLHNHYENSILRGV
LIDHPLSQADELVCTFSQVCRAVEIMOTLHPDASANFSLDLYVFGGONAIQI
AIYAHQPTRADEIPMBPBDILGVAGNHMDYSKGVNRLIGRTGLYPSYKVRREKIEYK
YPTYPEAEK"
2218
/gene="FUT8"
/note="11"
/frequency=".5"
/replace="C"
2758..2763
/gene="FUT8"
/note="putative"
2770..2775
/gene="FUT8"
/note="putative"
2869..2874
/gene="FUT8"
/note="putative"
polyA_signal
polyA_signal
polyA_signal
BASE COUNT 905 a 597 c 663 g 827 t
ORIGIN
Query Match 87.4%; Score 1510.4; DB 9; Length 2992;
Best Local Similarity 92.1%; Pred. No. 0;
Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
OY 1 ATCGGGCCATGAGACTGGTTCGTCGGCGTTGGATTATGCTATCTTTTGGCTGGGAGC 60
Db 429 ATCGGGCCATGAGACTGGTTCGTCGGCGTTGGATTATGCTATCTTTTGGCTGGGAGC 488
OY 61 TTCTATTTTACTAGTGTCTCTGCTGTTGTCGAGATATGACCACTGTACTCTAGC 120
Db 489 TTCTGTTTATATAGTGTCTCTGCTGTTGTCGAGATATGACCACTGTACTCTAGC 548
OY 121 CGAAGCTGTCGAAGATTTTGGCAAGAGCTGGAAGGCTTAAACACAAATGAAAGCTTG 180
Db 549 CGAAGCTGTCGAAGATTTTGGCAAGAGCTGGAAGGCTTAAACACAGATGAAAGCTTG 608
OY 181 AGGAGATGCGTGGATCTCTCCGAATACGAAAGGCCCATTTGATCAGGGCCAGCTTCA 240
Db 609 AGGAGATGCGCGAATCTCTCCGATACCGAAGGCCCATTTGATCAGGGCCAGCTATA 668
OY 241 GGAAGATTTGCTTTAAGAAGCAATTTATGAAGGCCAAAGACATTTGAAATTTAT 300
Db 669 GGAAGATTTGCTTTAAGAAGCAATTTATGAAGGCCAAAGACATTTGAAATTTAT 728
OY 301 AACAACCAACTAAATATGTTGCCAGGAAGAGATCATGAATCTTAAGAGAGAGATTTGAA 360
Db 729 AACAACCAAGCAAAATGTTGCGGGAAGAGATCATGAATCTTAAGAGAGAGATTTGAA 788
OY 361 AATGAGCTTAAGAGCTCTGCTTTTCTAACAAGTGAAGTGAAGAAATTTAAGAAATTTA 420
Db 789 AATGAGCTTAAGAGCTCTGCTTTTCTAACAAGTGAAGTGAAGAAATTTAAGAAATTTA 848
OY 421 GAAGGAATTAACCTCCAAAGACATGAGATGATTTTCTATCAATTTGGACATCATGAA 480
Db 849 GAAGGAATTAACCTCCAAAGACATGAGATGATTTTCTATCAATTTGGACATCATGAA 908
OY 481 AGGCTATATGACGGATCTATCTACCTAGTCAACACAGATGGGCGAGTGTGGCGT 540
Db 909 AGGCTATATGACGGATCTATCTACCTAGTCAACACAGATGGGCGAGTGTGGCGG 968
OY 541 GAAAGAGGCCAAGATCTGACAGAGCTGTCACGCGAAGTAACATATTTTCAGAAAT 600
Db 969 GAAAGAGGCCAAGATCTGACAGAGCTGTCACGCGAAGTAACATATTTTCAGAAAT 1028
OY 601 CCCAAGAGCTGCGACCAAGCAAGCTATGTTGATATTCACAAAGGCTGTGGCTAT 660
Db 1029 CCCAAGAGCTGCGACCAAGCAAGCTATGTTGATATTCACAAAGGCTGTGGCTAT 1088

QY 661 GGCTGACCTCCATCATGTAGTGTACTGCTTTATGATTCATATGACACCCAGCAGACA 720
|||||
DB 1089 GGCTGTGACCTCCATCATGTATGGTCTTACTCTTCATGATTCATATGGCACCAGCCAGACA 1148
QY 721 CTGCGCTTGGAAATCTCAAAATTTGGCGCTACGCTACGGGGGATGGGAACTGTATTAGA 780
||| |||||
DB 1149 CTATCATTTGGAAATCTCAAAATTTGGCGCTATGCTACTGGATGGAGAGACTGATTTAG 1208
QY 781 CCTGTAACTGACACGTGCACACAGATGTGGCAGCTCCACTGGACATTTGCTACAGTGA 840
|||||
DB 1209 CCTGTAACTGACACGTGCACACAGATGTGGCAGCTCCACTGGAGACTGCTAGAGTGA 1268
QY 841 GTAAAGGACAAAATTTGAGGTTGAGTCCCATTTGTAGACAGCTGATCCTCTCT 900
|| |||||
DB 1269 GTAAAGGACAAAATTTGAGGTTGAGTCCCATTTGTAGACAGCTGATCCTCTCT 1328
QY 901 CCTCATATTTTACCCCTGGCTGCTCCAGAAAGCTTGCAGATCGACTTGTAGAGTCCAT 960
|||||
DB 1329 CCTCATATTTTACCCCTGGCTGCTCCAGAAAGCTTGCAGATCGACTTGTAGAGTCCAT 1388
QY 961 GGTGATCTGCACTGTGGTGGTATCCCACTTTGTCAGTACTTGTATCCGCCACACACC 1020
|||||
DB 1389 GGTGACCCCTGCAGTGTGGTGGTGTCTCAGTTGTCAAAATCTGATCCGCCACACACC 1448
QY 1021 TGGCTGGAAGAAATAGAGAGGCCACCAAGAACCTAGGCTTCAACATCCAGTTATT 1080
|||||
DB 1449 TGGCTGGAAGAAATAGAGAGGCCACCAAGAACCTAGGCTTCAACATCCAGTTATT 1508
QY 1081 GGAGTCCATGTTAGACGACAGACAAAGTGGAGGCGGAGACGCTTCCATCCATTTGAG 1140
|||||
DB 1509 GGAGTCCATGTTAGACGACAGACAAAGTGGAGGAGACAGAGCTGCTTCCATCCATTTGA 1568
QY 1141 GAATACAGCGTGCAGCTTGAAGAGACTTTCAGCTTCTTGCTGCGAATGCAAGTGAAT 1200
|| |||||
DB 1569 GAGTACATGCTGCATGTTGAAGAACATTTTTCAGCTTCTTGCAAGTGAAGTGAAG 1628
QY 1201 AAAAAAGGCTGATTTGGCCACAGATGACCCCTTGTAAAGAGGCAAAAACAAG 1260
|||||
DB 1629 AAAAAAGGCTGATTTGGCCACAGATGACCCCTTGTAAAGAGGCAAAAACAAG 1688
QY 1261 TACCCAGTATGAAATTTATAGTGAATCTATCTTTGGTGCAGCTGAGTACATTAAT 1320
|||||
DB 1689 TACCCAGTATGAAATTTATAGTGAATCTATCTTTGGTGCAGCTGAGTACATTAAT 1748
QY 1321 CGATATACGAAAAATTCACCTTGGGGTGTGATCTGGATATACACTTCTCTCCAGGCA 1380
|||||
DB 1749 CGATATACGAAAAATTCACCTTGGGGTGTGATCTGGATATACACTTCTCTCCAGGCA 1808
QY 1381 GACTTCTGATGTGATCTTTTTCATGCGAGTCTGTAGAGTTCCTATGAATCATGCA 1440
|||||
DB 1809 GACTTCTGATGTGATCTTTTTCATGCGAGTCTGTAGAGTTCCTATGAATCATGCA 1868
QY 1441 GGCTGTCATCTGATGCTCTGCGAATCTCCGCTCTTGGATGACATCTACTATTTTGA 1500
|| |||||
DB 1869 ACATGTCATCTGATGCTCTGCGAATCTCTGCAATCTTTAGATGACATCTACTATTTTGG 1928
QY 1501 GGCCCAATGCCCAACCAAAATTTGCCATTTATCTCACCAACCTTGAACCTGAAGAGAA 1560
|||||
DB 1929 GGCCCAATGCCCAACCAAAATTTGCCATTTATGCTCACCAACCTTGAACCTGAAGAGAA 1988
QY 1561 ATCCCATGGAAGCTGGAGATTTATTTGGTGGCTGCAAAATCATCTGGATGGCTATCTCT 1620
|| |||||
DB 1989 ATTTCCCATGGAAGCTGGAGATTTATTTGGTGGCTGCAAAATCATCTGGATGGCTATCTCT 2048
QY 1621 AAAGTGTTTACGAAAACTGGGAAGAGCGGCTTATTCCTCTTACAAAGTTTCAGAG 1680
|||||
DB 2049 AAAGTGTTTACGAAAACTGGGAAGAGCGGCTTATTCCTCTTACAAAGTTTCAGAG 2108
QY 1681 AAGATGAAACAGTCAAGTACCCACATATCCCGAGGCTGACAAAGTAA 1728
|||||
DB 2109 AAGATGAAACAGTCAAGTACCCACATATCCCGAGGCTGACAAAGTAA 2156

RESULT 9
HSY17978
LOCUS 3186 bp mRNA linear PRI 06-NOV-1998
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase
transcript A1.
ACCESSION Y17978
VERSION Y17978.1 GI:3451266
KEYWORDS FUT8 gene; glycoprotein 6-alpha-L-fucosyltransferase.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE
AUTHORS Callieau,A., Balanzino,L., Candelier,J.J., Oriol,R. and Mollicone,R.
TITLE Differential splice variants of human FUT8 embryonic cDNA
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3186)
AUTHORS Callieau,A.L.V.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1998) A.L.V. Callieau, INSERM U178, 16 av P.V. Couteurier, 94807 Villejuif Cedex, FRANCE
COMMENT Related entry: D89289.
FEATURES
source location/Qualifiers
1..3186
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q23"
/clone="FUT8 transcript A1"
/tissue_type="whole embryo"
/dev_stage="40-70 days"
1..3162
/gene="FUT8"
1..389
/gene="FUT8"
/note="A"
390..488
/gene="FUT8"
/note="C"
489..2444
/gene="FUT8"
/note="D"
717..2444
/gene="FUT8"
/EC_number="2.4.1.68"
/note="allele 1"
/codon_start=1
/product="glycoprotein 6-alpha-L-fucosyltransferase"
/protein_id="CA76987.1"
/db_xref="GI:3451267"
/translation="MRPMTGSWRWIMLIFAFGTLTFYTGHLVRDNDHPDISSNELS
KTLAKLERKQONEDLRMAESLRIPBGRIDGPAIGRVRLERDLVAKKIDLEVRK
OTRNGLGDHETLRRIENGAKELNFPQSELSKKLNLEGNLORNADEPLDLDGHE
RSIMTDLTYSQDAGDMREKADLELVORITVYLDNPKDSKAKLVGNINKG
GEGCOLHHVVCYFMFLAYGTQRTLIESONMVRATGWEVFPVSECTKDVSNISG
WSGEVKDRNVQVVELPIYDSLHPRPYPLAVPEDLADRLVRHGDPAVMVVSQFYK
LIRPQMLEKEIEEATKKLGFKHPVIGVHVRTRKDVGEAFAHPLEEVYVHYEHFOL
LARMQVKKRYVLAITDDPSLLEAKETKYPNPEFTSDMSISGALHNRYPENSLRGV
LIDHIFLSQADPLVCTSPSOVCRAVEYETKOTLHPDASANFHSIDITYPCGONAHNOI
AIVAHQPRADEIPMEPDITIGVAGNHWDGYSKGNRRKLGRTGLPSYKVRKEIETVK
IPTYPEAK"

variation

/gene="FUT8"
/note="I"
/frequency=".5"
/replacem="c"
/gene="FUT8"
/note="putative"
3058..3063
polyA_signal
polyA_signal
polyA_signal

/gene="FUR8"
/note="putative"
3157..3162
polyA_signal
/gene="FUR8"
/note="putative"
BASE COUNT 910 a 710 c 721 g 840 t 5 others
ORIGIN
Query Match 87.4% Score 1510.4 DB 9 Length 3186;
Best Local Similarity 92.1% Pred. No. 0;
Matches 1592: Conservative 0; Mismatches 136; Indels 0; Gaps 0;
OY 1 ATGCGGCATGAGCTGTTGCTGCGCTGGATTAATGCTCATCTTTTGGCTGGGACC 60
Db ATCGGCGCATGAGCTGTTGCTGCGCTGGATTAATGCTCATCTTTTGGCTGGGACC 716
OY 61 TTGCTATTTTACATAGTGTCTGTTGTAACGATTAATGACACCTGTACTCTAGC 120
Db TTGCTATTTTATATAGTGTCTGTTGTAACGATTAATGACACCTGTACTCTAGC 836
OY 121 CGAGACTGTCAGATTTTGGCAAGCTGGAACGCTTAAACAAATAAGAACTTG 180
Db CGAGAACTGTCAGATTTTGGCAAGCTTGAACGCTTAAACAAATAAGAACTTG 836
OY 181 AGAGAGATGCTGATCTCTCGAATACGAAAGGCCCTTGTATGAGGGCCAGCTTCA 240
Db AGCGAATGGCCGATCTCTCGGATACGAAAGGCCCTTGTATGAGGGCCAGCTTCA 897
OY 241 GGAAGCTTCGCTTTAGAGAGCAATTTATGAGGCCAAAGAACAGATTTAAATTAT 300
Db GGAAGAGTCCGCTTTAGAGAGCAATTTATGAGGCCAAAGAACAGATTTAAATTAT 957
OY 301 AAGAAACAACCTAAATGGTCCAGGGAAGGATCATGAATCCTAAGGAGGATTTGA 360
Db AAGAAACAACCAAGATGGTCTGGGAGAGATCATGAATCCTGAGAGGAGATTTGA 1017
OY 361 AATGAGCTAAAGAGCTGTGTTTTTCTACAAAGTGAAGTTGAAGAAATTTAA 420
Db AATGAGCTAAAGAGCTGTGTTTTTCTACAAAGTGAAGTTGAAGAAATTTAA 1077
OY 421 GAAGGAATGAACTCCAAAGACATGAGATTTCTATCAGATTTGGGACATCTGAA 480
Db GAAGGAATGAACTCCAAAGACATGAGATTTCTATGAGATTTGAGATTTGAA 1137
OY 481 AGGTCTATATAGCGGATCTATCTACTCAGTCAAAAGATGGGAGGATTTGGCGT 540
Db AGGTCTATATAGCGGATCTATCTACTCAGTCAAAAGATGGGAGGATTTGGCGG 1197
OY 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCCAGCGGAATTAATATCTTCAGAA 600
Db GAAAGAGAGCCAAAGATCTGACAGAGCTGTCCAGCGGAATTAATATCTTCAGAA 1257
OY 601 CCCAAGGATCGCAAGAGCCAAAGCTAGTGTATATATCAAAAGGCTTGGCTAT 660
Db CCCAAGGATCGCAAGAGCCAAAGCTAGTGTATATATCAAAAGGCTTGGCTAT 1317
OY 661 GGCTGTACAGCTCATATGATGATCTATGATTTGATTTGACACCCAGCAACA 720
Db GGCTGTACAGCTCATATGATGATGATCTATGATTTGATTTGACACCCAGCAACA 1377
OY 721 CTCGCTTGGAAATCTCAGAAATGGGCGCTACGCTACTGGGGGATGGAAACTGT 780
Db CTCGCTTGGAAATCTCAGAAATGGGCGCTATGCTACTGGGAGATGGGAACTGT 1437
OY 781 CCTGTAAAGTGAAGCTGACAGACAGATCTGGGAGCTCCACTGGAATTTGGTGA 840
Db CCTGTAAAGTGAAGCTGACAGACAGATCTGGGAGCTCCACTGGAATTTGGTGA 1497
OY 841 GTAAGGACAAAATTTTCAAGGTGTTGAGCTCCCATTTGTACAGTGTTCCTCGT 900
Db GTAAGGACAAAATTTTCAAGGTGTTGAGCTCCCATTTGTACAGTGTTCCTCGT 1557
OY 901 CCTCATATTTTACCCCTGTGCCAGAAAGACCTTGCAGATGCACTTGTACAGTCA 960

||||| 1617 CTTCCATATTTTACCTTGGCTGTACCAAGAAAGACCTGCGAGATGACTTGTACAGTGCAT 1676
OY 961 GGTGATCTCGCATGTTGGTGGTATCCAGTTGTGTCAACTGTGTTGGCCACAGACC 1020
Db 1677 GGTGACCTTCAGTGTGGTGGTGTCTCAATTAATCTGTGATCCGCCACAGCCT 1736
OY 1021 TGGCTGGAAGAAATAGAGAGAGCCACCAAGAGACCTAGCTTCAACATCCAGTTAT 1080
Db 1737 TGGCTGGAAGAAATAGAGAGAGCCACCAAGAGACCTAGCTTCAACATCCAGTTAT 1796
OY 1081 GGAATCCATTTAGACGACACAGCAAAAGTGGAGGCGGAGACCTTCATCCATTGAG 1140
Db 1797 GGAATCCATTTAGACGACACAGCAAAAGTGGAGGCGGAGACCTTCATCCATTGAG 1856
OY 1141 GATPACAGGTGACGTGTGAAGAACTTACCTTCTGCTGGCAATTCGAAGTGGAT 1200
Db 1857 GATPACAGGTGACGTGTGAAGAACTTACCTTCTGCTGGCAATTCGAAGTGGAT 1916
OY 1201 AAAAAAGGCTGATTTGGCCACAGATGACCCCTGTTGTTAAAGAGCCAAACAAAG 1260
Db 1917 AAAAAAGGCTGATTTGGCCACAGATGACCCCTGTTTAAAGAGGCAAAACAAAG 1976
OY 1261 TACCCAGTTATGATTTATTTAGTATTAACCTATCTTGTGCTGACCTGACTATTAAT 1320
Db 1977 TACCCAGTTATGATTTATTTAGTATTAACCTATCTTGTGCTGACCTGACTATTAAT 2036
OY 1321 CGATPACAGAAATTCACCTTGGGGTGTATCCCGGATATACCTTCTCCCGCAGGCA 1380
Db 2037 CGATPACAGAAATTCACCTTGGGGTGTATCCCGGATATACCTTCTCCCGCAGGCA 2096
OY 1381 GACTTCTAGTGTGATCTTTTTCATGCGAGTGTGAGATTTGCTTATGAATCATGCA 1440
Db 2097 GACTTCTAGTGTGATCTTTTTCATGCGAGTGTGAGATTTGCTTATGAATCATGCA 2156
OY 1441 GCGCTCATCTGATGCTCTCTGCAACTTCCGTTCTTGGATGACATCTACTATTTTGA 1500
Db 2157 ACATCATCTCTGATGCTCTCTGCAACTTCCATCTTGTAGATGACATCTACTATTTGGA 2216
OY 1501 GGCCCAATGGCCACACCAAAATTTGCAATTTATCCACCAACCTGGAAGAGAGAA 1560
Db 2217 GGCCCAATGGCCACACCAAAATTTGCAATTTATGCTCCACCAACCTGGAAGAGAA 2276
OY 1561 ATCCCATGGAACCTGAGATTTATTTGGTGTGCTGGAATCACTGGGATGCTATCTC 1620
Db 2277 ATCCCATGGAACCTGAGATTTATTTGGTGTGCTGGAATCACTGGGATGCTATCTC 2336
OY 1621 AAAGTGTTAACAGAAATCTGGGAGAGCGGCTTATTCCTCTTACAAAGTTCCAGAG 1680
Db 2337 AAAGTGTTCACAGAAATCTGGGAGAGCGGCTTATTCCTCTTACAAAGTTCCAGAG 2396
OY 1681 AAGATGAAGAACGTCAAGTACCCACATATCCGAGGCTGACAGATTA 1728
Db 2397 AAGATGAAGAACGTCAAGTACCCACATATCTGAGGCTGACAGATTA 2444
RESULT 10
HSY17979 3280 bp mRNA linear PRI 06-NOV-1998
LOCUS HSY17979
DEFINITION Homo sapiens mRNA for glycoprotein 6-alpha-L-fucosyltransferase
transcript A2.
ACCESSION Y17979.1 GI:3451268
VERSION Y17979.1
KEYWORDS FUR8 gene; glycoprotein 6-alpha-L-fucosyltransferase.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 3280)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS Calleeu, A., Balanzino, L., Candeller, J.J., Ortol, R. and
Mollicone, R.
TITLE Differential splice variants of human FUR8 embryonic cDNA
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 3280)
AUTHORS Cailliau,A.L.V.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-1998) A.L.V. Cailliau, INSERM U178, 16 av P.V.
Coururier, 94807 Villejuif Cedex, FRANCE
COMMENT Related entry: D89289.
FEATURES
source 1. 3280
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q23"
/clone="FUT8 transcript A2"
/tissue_type="whole embryo"
/dev_stage="40-70 days"
1. 3162
/gene="FUT8"
1. 389
/gene="FUT8"
/note="A"
390. 488
/note="C"
/gene="FUT8"
489. .>2444
/note="C"
/gene="FUT8"
/note="D"
717. 2444
/gene="FUT8"
/EC_number="2.4.1.68"
/note="allele I"
/product="glycoprotein 6-alpha-L-fucosyltransferase"
/protein_id="CAA76988.1"
/db_xref="GI:3451269"
/db_xref="SPTREMBL:O00235"
/translation="MRPTGSWRIMLILFANGTLLEYIGHLVBDNDHPDHSRELS
KILAKLERKQONEDLRMAESLRIPEDPIDGPAIGRVRLBQLVAKKEIDENYK
OTRNLGDHEILRRRIENGAKELMFEFDOSEKLIKLEGNLOHRADPEFLDLGHE
RSINTDLVYLSOTDAGDKREKBAKDLTELVDORITTYLDNPKDCSAKLVGNINCK
GYCOLHNVYCFMIAYGTQRTLLESOMRYATGMEVTFRPSVETCTDRSGISTGH
MSGEVDEKNVQVLEPIYDSLHPRPYLPYLAVEDLADLVVHGDPAVMWYSEVYK
LIRPOMLEKEIEEATKLGFKHPVIGVHVRDVKVTEAAPHIEEYVHVEEHOL
LARMOVKKRYVLATDDPSILKEAKTKYPNEYISDINSAGIHNYTPNSLRGY
LIDHIFLSADPLVCTFSSQYCRVAYELMOTLHPDASNFHSLDIIYYGGMANHOI
ATYAHPTADEIPMEPGDIGVAGNHMDYSKGVNRKLGRTGLPYKVKREKIEYK
LYYPEAEK"
2506
variation
/gene="FUT8"
/note="I1"
/frequency=".5"
/replace="C"
3046. 3051
/note="FUT8"
/gene="putative"
3058. 3063
/note="FUT8"
/note="putative"
3157. 3162
/gene="FUT8"
/note="putative"
BASE COUNT 934 a 718 c 743 g 880 t 5 others
ORIGIN
Query Match 87.4%; Score 1510.4; DB 9; Length 3280;
Best local Similarity 92.1%; Pred. No. 0;
Matches 1592; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
OY 1 ATCGGCGCATGAGCTGGTTCGGGTGATGATGCTCACTTTTGGCTGGGGACC 60
DB 717 ATCGGCGCATGAGCTGGTTCGGGTGATGATGCTCACTTTTGGCTGGGGACC 776
OY 61 TTGCTATTTTACATAGGTGCTCACTTGTACGAGATATGACCACCTGTGACTCTAGC 120
TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT TTTTT

DB 777 TTGCTGTTTTATATAGGTGCTCACTTGTATGAGATATGACCACCTGTACCTCTAGC 836
OY 121 CGAGAACTGTCCTCAAGATTTTGGCAAAAGCTGGAAACCTTTAAACAAATGAAGCTTG 180
TT
DB 837 CGGAACTGTCCTCAAGATTTTGGCAAAAGCTGGAAACCTTTAAACAAATGAAGCTTG 896
OY 181 AGGAGAACTGCTGATCTCTCCGATATACAGAAAGGCCCCATTTATAGGGGCCCTTCA 240
TT
DB 897 AGGCAATGCGCCGAATCTCTCCGATATACAGAAAGGCCCCATTTATAGGGGCCCTTCA 956
OY 241 GGAAGAGTTCTGCTTTTGAAGAGCAATTTATGAAGGCCCAAGAACAGATTGAAATTTAT 300
TT
DB 957 GGAAGAGTTACGGCTTTTGAAGAGCAAGCTTTGTTAAGGCCAAGAACAGATTGAAATTTAC 1016
OY 301 AAGAACAACTAATAAATGCTCCAGGAAGAGATCATGAATCTTAAGAGAGAGATTGAA 360
TT
DB 1017 AAGAACAGACCCAGAAATGCTCGGGGAAGATCATGAATCTTGAGGAGAGATTGAA 1076
OY 361 AATGAGCTAAAGAGCTCGGTTTCTTACAAAGTGAAGTGAAGAAATTTAAGAAATTTTA 420
TT
DB 1077 AATGAGCTAAAGAGCTCTGGTTTTTCTTACAGAGTGAATTTGAAGAAATTTAAGAACTTTA 1136
OY 421 GAAGAAATGAACCTCCAAAGACATGAGATGAATTTCTATCAGATTTGGACATCATGA 480
TT
DB 1137 GAAGAAATGAACCTCCAAAGACATGAGATGAATTTCTTTTGGATTTAGACATCATGA 1196
OY 481 AGGCTATTAATGACGATCTATCTACCTCAGTCAAGACAGATGGGAGCTGATTTGGCGT 540
TT
DB 1197 AGGCTATTAATGACGATCTATCTACCTCAGTCAAGACAGATGGGAGCTGATTTGGCGG 1256
OY 541 GAAAGAGAGCCCAAGATCTGACAGAGCTGTCACAGGGAATTAATATCTTTCAGAA 600
TT
DB 1257 GAAAGAGAGCCCAAGATCTGACAGAGCTGTCACAGGGAATTAATATCTTTCAGAA 1316
OY 601 CCCAAGAGCTGACAGCAAGGCCAAGAGCTAGTGTGAATTTCAACAAGAGCTGTGGCTAT 660
TT
DB 1317 CCCAAGAGCTGACAGCAAGGCCAAGAGCTAGTGTGAATTTCAACAAGAGCTGTGGCTAT 1376
OY 661 GGCTGTACGCTCATCATCTAGTGTACTGCTTTATGATTTGCATATGGCACCCAGCAACA 720
TT
DB 1377 GGCTGTACGCTCATCATCTAGTGTACTGCTTTATGATTTGCATATGGCACCCAGCAACA 1436
OY 721 CTCGCCCTGGAATCTCAATTTGGCCCTACGCTACGCTGGGGAGAGGAAACTGTGTTAGA 780
TT
DB 1437 CTCATCTTGAATCTCAATTTGGCCCTATGCTATGCTAGTGGAGAGGAGCTGTGATTTAGG 1496
OY 781 CCTGTAAAGTGAAGAGCTGACAGACAGATCTGCGACCTCCACTGAGACTTGTGACGTGA 840
TT
DB 1497 CCTGTAAAGTGAAGAGCTGACAGACAGATCTGCGACCTCCACTGAGACTTGTGACGTGA 1556
OY 841 GTAAAGACAAAATGTTTCAGGTGTTGAGTCTCCCATTTGTACAGAGTGTTCATCTCGT 900
TT
DB 1557 GTAAAGACAAAATGTTTCAGGTGTTGAGTCTCCCATTTGTACAGAGTGTTCATCTCGC 1616
OY 901 CCTGCATATTTACCCCGGCTGTCAGAAAGACCTTGAGATGACACTTGTAGAGAGTTCAT 960
TT
DB 1617 CCTGCATATTTACCCCGGCTGTCAGAAAGACCTTGAGATGACACTTGTAGAGAGTTCAT 1676
OY 961 GGTGATCTGTCAGTGTGTTGGGTATCCCACTTTGTCAAGTACTTGTATTCGCCCAACACC 1020
TT
DB 1677 GGTGATCTGTCAGTGTGTTGGGTATCCCACTTTGTCAAGTACTTGTATTCGCCCAACACC 1136
OY 1021 TTGCTGAAAAGGAATATGAAAGAGCCACCAAGAGCTAGCTTTCAAACTTCAGTTATT 1080
TT
DB 1737 TTGCTGAAAAGGAATATGAAAGAGCCACCAAGAGCTAGCTTTCAAACTTCAGTTATT 1796
OY 1081 GGAATCCATGTTAGAGCAGACAGAAAGTGGAGGGAAGCAGCTTCATCCCATTTGAG 1140
TT
DB 1797 GGAATCCATGTTAGAGCAGACAGAAAGTGGAGGGAAGCAGCTTCATCCCATTTGAG 1856
OY 1141 GAATACAGCGTGCACGTTGAAGAAGACTTTCACCTTCTGCGAGAAATGCAAGTGGAT 1200
TT
DB 1857 GAGTACATGCTGATGTTGAAGAACAATTTTCACCTTCTGCGAGAAATGCAAGTGGAC 1916

QY	1261	TACCCAGTATGAATTTAGTATGATACCTATCTCTGGTCAGCTGACATATAT	1320
DB	1458	TACCCCAATTTATGATTTATATGATTAACCTATTTCTGGTCAGCTGACATAT	1517
QY	1321	CGATATACAGAAATTTCACTTGGGGTGTGATCTGATATACATTTCTCTCCAGCA	1380
DB	1518	CGATACAGAAATTTCACTTGGGGTGTGATCTGATATACATTTCTCTCCAGCA	1577
QY	1381	GACTCTAGTGTGATCTTTTATGATGAGCTGTATGATGCTTATGATATCATGCA	1440
DB	1578	GACTCTAGTGTGATCTTTTATGATGAGCTGTATGATGCTTATGATATCATGCA	1637
QY	1441	GCGTGCATCTGATGCTCTGCAACTCTGCTCTTTGATGATCATATTTTGA	1500
DB	1638	ACACTACATCTGATGCTCTGCAACTCTGCTCTTTGATGATCATATTTTGA	1697
QY	1501	GCCCCAAATGCCACACCAATTTGCATTTATCTCCACCACTGCACTGAGAGAA	1560
DB	1698	GCCCCAAATGCCACACCAATTTGCATTTATCTCCACCACTGCACTGAGAGAA	1757
QY	1561	ATCCCCAGAACTGAGATATTTATGCTGTGCTGGAATCACTGGGATGCTATCT	1620
DB	1758	ATCCCCAGAACTGAGATATTTATGCTGTGCTGGAATCACTGGGATGCTATCT	1817
QY	1621	AAAGTGTTAACAGAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1680
DB	1818	AAAGTGTTCACAGAAATTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1877
QY	1661	AAAGTGAAGACAGTCACTACCCATATATCCGAGGCTGACAGTAA	1728
DB	1878	AAAGTGAAGACAGTCACTACCCATATATCCGAGGCTGACAGTAA	1925

RESULT 12
BC010666 .2976 bp mRNA linear ROD 07-AUG-2002

LOCUS Mus musculus, fucosyltransferase 8, clone MGC:11418 IMAGE:3594582,
DEFINITION mRNA, complete cds.
ACCESSION BC010666
VERSION BC010666.1 GI:14715012
KEYWORDS MGC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2976)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amgdbcm.tmc.edu
Günatane, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

FEATURES
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAC Plate: 16 Row: 1 Column: 5
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8393370.
Location/Qualifiers

		source	
		1. .2976 /organism="Mus musculus" /db_xref="taxon:10090" /map="FVB/N" /clone="MGC:11418 IMAGE:3594582" /lssue-type="Mammalian tumor, C(3)1-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse." /clone_1db="NCI_CGAP_Mam6" /1db_host="DH10B" /note="Vector: pCMV-SPORT6" 584. .2311 /codon_start=1 /product="fucosyltransferase 8" /protein_id="AAH10666.1" /db_xref="GI:14715013" /db_xref="LocusID:53618" /translation="MRAVTCGSMRWIMLIFAMGTLIFYGHLVRDNDHPHSSRELS KILAKLERKQNDLRRMARESLRPEPIQDQATGVRVLEOLVAKAEIOENYK QARNGICQDEILRRRIENKAKELMFLPSGLKKLKLKEGNEILRHADKLLDNGH RSIMTDLYLLSQTGAGDMREKADLELVQRRTYTLQNRKDCSKARKLYCNKGC GYGCOLHNVICFMIVATGRTLLLESOMRYATGMEVFRVSECTDQSGLSTGH WSGEVNDKNIQVLELIVDSLHPPRYPLAIVPDLRLRVHGDPAVWVVSQVRY LIRQPMLEKEIEBETKLGFKHPTGVHVRTRKVGTEAALFHP1 EYEMVHVEHFO LARRMVDKRRVYVATDPTLLKEAKTKYSYERISDSISMSAGLHNRRTENSILRG ILDIHPLSQADEFLVCTSSQVCRVAEIMOTLHPDASNFSLDDIYFEGOMAHNOI AVYPHKPRTEIEITMERODITGVAGNHMDGTSKGINRKLGTGLYPSIKVREKLETYK YPTYPEAK"	
BASE COUNT		909 a 626 c 660 g 781 t	
ORIGIN			
Query Match		84.9% Score 1467.8; DB 10; Length 2976;	
Best Local Similarity		90.6% Pred. No. 0;	
Matches 1565; Conservative		0; Mismatches 162; Indels 0; Gaps 0;	
QY	1	ATCGCGCATGAGCTGCTGCTGCTGGATTAGCTCATCTTTTGGCTGGGGAGC	60
DB	584	ATCGCGCATGAGCTGCTGCTGCTGGATTAGCTCATCTTTTGGCTGGGGAGC	643
QY	61	TTGCTATTTTACATAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	120
DB	644	TTGCTATTTTACATAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	703
QY	121	CGAGAGCTGCTCAAGATTTTGGCAAGCTGCAAGCTTAAACCAACCAATTAAC	180
DB	704	AGAGAGCTGCTCAAGATTTTGGCAAGCTGCAAGCTTAAACCAACCAATTAAC	763
QY	181	AGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	240
DB	764	AGGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	823
QY	241	GGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	300
DB	824	GGAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	883
QY	301	AGGAACAACCTAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	360
DB	884	AGGAACAACCTAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	943
QY	361	AATGAGCTAAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	420
DB	944	AATGAGCTAAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1003
QY	421	GAGGAAATGAATCTCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	480
DB	1004	GAGGAAATGAATCTCAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	1063
QY	481	AGCTCTAATGAGAGCTTATACCTACCTACCTACCTACCTACCTACCTACCT	540
DB	1064	AGCTCTAATGAGAGCTTATACCTACCTACCTACCTACCTACCTACCTACCT	1123
QY	541	GAAAGAGAGGCTCAAGATCTGACAGAGCTGCTGCTGCTGCTGCTGCTGCT	600
DB	1124	GAAAGAGAGGCTCAAGATCTGACAGAGCTGCTGCTGCTGCTGCTGCTGCT	1183


```

1 Db 385 AGCGAATGGCTGAGTCTCTCCGAATACCGAAGGCCCATTTGACGAGGGGACAGCTACA 444
Oy 241 GGAAGAGTTCGTCCTTTAGAGAGCAATTTATGAAGGCCCAAGACAGATTTGAAAATTTAT 300
Db 445 GGAAGAGTCCGCTGTTTACAGAGACAGCTTTGTAAGGCCCAAGAACAGATTTAAAATTTATC 504
Oy 301 AAGAAACAACACTAAAATATGTCAGAGAGATTCATGAATCTTAAGAGAGATTTGA 360
Db 509 AAGAAACAACAGTGAATATGTCAGAGAGATTCATGAATCTTTAAGAAGAGATTTGA 564
Oy 361 AATGAGCTAAAGAGCTGCTGTTTTTTCTACAAGTGAAGTGAAGAAATTTAAGAATTTTA 420
Db 565 AATGAGCTAAAGAGCTGCTGTTTTTTCTACAAGTGAAGTGAAGAAATTTAAGAATTTTA 624
Oy 421 GAAGAAATGAAGTCCAAAGACATGAGTGAATTTCTATCATGATTTGGACATCATGAA 480
Db 625 GAAGAAATGAAGTCCAAAGACATGAGTGAATTTCTTGTGATTTAGACACCATGAA 684
Oy 481 AGGTCTATATGACGAGATCTATCTACCTCAAGTCAAGACATGAGGAGGAGATTTGGCT 540
Db 685 AGGTCTATATGACGAGATCTATCTACCTCAAGTCAAGACATGAGGAGGAGATTTGGCT 744
Oy 541 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCACAGGAGAAATACATATCTTCAGAA 600
Db 745 GAAAGAGAGCCAAAGATCTGACAGAGCTGTCACAGGAGAAATACATATCTTCAGAA 804
Oy 601 CCCAAGAGCTGACGAGAAAGCCAGAAAGCTAGTGTAAATATACAAAGCTGTGGCTAT 660
Db 805 CTTAAGAGCTGACGAGAAAGCCAGAAAGCTAGTGTAAATATACAAAGCTGTGGCTAT 864
Oy 661 GGCTGTCAAGCTCCATCATGATGATGATCTTATGATGATGATGATGATGATGATGATGAT 720
Db 865 GGCTGTCAAGCTCCATCATGATGATGATGATCTTATGATGATGATGATGATGATGATGATGAT 924
Oy 721 CTCGCTTGAAGTCTCAGCAATGAGGCTACGCTACTGAGGAGTGGGAACTGTGTTTGA 780
Db 925 CTCGCTTGAAGTCTCAGCAATGAGGCTATGCTACTGAGGAGTGGGAACTGTGTTTGA 984
Oy 781 CCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 840
Db 985 CCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1044
Oy 841 GTAAAGAGCAAAATGTCAGTGTGTTGAGTCCCATTTGATGATGATGATGATGATGATGATGAT 900
Db 1045 GTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1104
Oy 901 CCTCATATATTTACCCCTGCTGCCAGAGACCTTGACGATGATGATGATGATGATGATGATGAT 960
Db 1105 CCTCATATATTTACCCCTGCTGCCAGAGACCTTGACGATGATGATGATGATGATGATGATGAT 1164
Oy 961 GGTGATCCGTCAGTGTGTTGGTATCCCATTTGTCATGATGATGATGATGATGATGATGATGAT 1020
Db 1165 GGTGATCCGTCAGTGTGTTGGTATCCCATTTGTCATGATGATGATGATGATGATGATGATGAT 1224
Oy 1021 TGGCTGAAAAGAGAAATAGAGAGGACCAAGAGCTAGCTTCAACATCCAGTATTT 1080
Db 1225 TGGCTGAAAAGAGAAATAGAGAGGACCAAGAGCTAGCTTCAACATCCAGTATTT 1284
Oy 1081 GGAATTCATGTTAGAGCAGACAGAAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
Db 1285 GGAATTCATGTTAGAGCAGACAGAAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1344
Oy 1141 GAATTCAGGTCGACGTTGAAGAGACTTTCAGCTTCTTCGAGAGATGAGAGATGAGAT 1200
Db 1345 GAGTCAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1404
Oy 1201 AAAAAAGAGGTCATTTGGCCACAGATGACCTGCTTTTAAAGAGGCAAAAAACAAG 1260
Db 1405 AAAAAAGAGGTCATTTGGCCACAGATGACCTGCTTTTAAAGAGGCAAAAAACAAG 1464
Oy 1261 TACCCAGTATGAATTTATTTAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 1320
Db 1465 TACCCAGTATGAATTTATTTAGTGAATGATGATGATGATGATGATGATGATGATGATGAT 1524

```

```

Oy 1321 CGATATACAGAAAATTTACTTGGGGTGTGATCTCGGATATACACTTCTCTCCAGCA 1380
Db 1525 CGGTACACAGAAAATTTACTTGGGGTGTGATCTCGGATATACACTTCTCTCCAGGCT 1584
Oy 1381 GACTTCTAGTGTGATCTTATTTTATGAGAGTCTGATGATGATGATGATGATGATGATGATGAT 1440
Db 1585 GACTTCTAGTGTGATCTTATTTTATGAGAGTCTGATGATGATGATGATGATGATGATGATGAT 1644
Oy 1441 GGGCTGATCTGATGATCTCTGCGAAGCTTCCGTTCTTTGATGATGATGATGATGATGATGAT 1500
Db 1645 ACCCTGATCTGATGATCTCTGCGAAGCTTCCGTTCTTTGATGATGATGATGATGATGATGAT 1704
Oy 1501 GGGCCAAATGCCCCAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db 1705 GGGCCAAATGCCCCAACAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1764
Oy 1561 ATCCCATGGAACCTGAGATATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
Db 1765 ATTCATGGAACCTGAGATATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1824
Oy 1621 AAGGTGTTACAGAAAATCTGGAGAGCGGCTATATCTCTTACAAAGTTCGAGAG 1680
Db 1825 AAGGTATCAACAGAAAATCTGGAGAGCGGCTATATCTCTTACAAAGTTCGAGAG 1884
Oy 1681 AAGATGAACAGTCAATATCCCATATCTCCGAGGCTGACAGTA 1727
Db 1885 AAGATGAACAGTCAATATCCCATATCTCCGAGGCTGACAGTA 1931

RESULT 14
BC025385
LOCUS BC025385 2682 bp mRNA linear PRI 07-AUG-2002
DEFINITION Homo sapiens, clone MGC:26465 IMAGE:4838097, mRNA, complete cds.
ACCESSION BC025385
VERSION BC025385.1 GI:22135669
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2682)
AUTHORS Strausberg,R.
TITLE Direct Submision
JOURNAL Submitted (05-MAR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA.
REMARK
NIH-MGC Project URL: http://mgc.ncl.nih.gov
COMMENT
Contact: MGC help desk
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiroki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL bc: http://image.llnl.gov
Series: IRK Plate: 34 Row: n Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4758407.
FEATURES
source
1..2682
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:26465 IMAGE:4838097"
/tissue_type="testis"
/clone_1ib="NIH_MGC_97"

```

```

/lab_host="DH10B"
/note="Vector: pBluescript"
597. .1937
CDS

```

```

/product="Unknown (protein for MGC:26465)"
/protein_id="AAH23385.1"
/db_xref="gi:22135670"
/translation="MATTYSLVNNKKRIIVLAQPTTVYKRRKIPQLKSIIMDLVYLSQTDGAGSRKEAKDLTEVQRITVYLDQPKDSKAKLVCNINCKGCGCOLHAYVEFMIAVGTQRTLLIESQNMPLRYATGMEVETPRVSPCTSPRSISTHMSGSEKQKVVAVLEPPYDLSHPRPPIPLVPEEDLADLVLRVNGDPAVVMVVSQFVYLLIRPOYMLEKLEIEATKILGRKHPVIGVHVRTDKVTEAFHPHIEEYMAVHEEFLRLRQVQDKRRVYLATDPSLSLKAETKRYPMVEFISDNISMSGGLNNRYTENSLLVLDIHFHLSADELVCOTSSQCRVAYELMOTLHPDASNFSPISLDIIVFGGQNNHNDIIVAIHQPRADELPMERPDITGAVGNHDDGYSKGVNKKLGTGLPYSKVREKLEIYKIPPTYEAKR"

```

BASE COUNT	808 a	554 c	597 g	723 t
ORIGIN				

Query Match	70.38;	Score 1215.4;	DB 9;	Length 2682;
Best Local Similarity	86.08;	Pred. No. 0;		
Matches 1403; Conservative	0;	Mismatches 126;	Indels 102;	Gaps 1;

OY	200	TCGGAATACGAAAGGCCCCATTGATTCAGGGGCCAGCTTCAGGAAGAGTCCGCTTTAG	259
Db	307	TCAGGATACCGAAAGGCCCTATTGTCAGGGGCCAGCTATAGGAAGAGTACCGCTTTAG	366
OY	260	AAGACCAATTTATGAAAGGCCAAAGAACAGATTGTGAATAATTTAAGAAACAATATAAAATG	319
Db	367	AAGACGACGTTGTTAAGGCCAAAGAACAGATTGAAATAATTTACAAAGAACAGACAGAAATG	426
OY	320	GTCCAGGGGAGGATCATGAAATCCTTAAGGAGAGGATTTGAAAATTTGAGCTAAAGAGCT	379
Db	427	GTCTGGGGAGGATCATGAAATCCTGAGAGAGGAGATTGAAAATTTGAGCTAAAGAGCT	486
OY	380	GGTTTTTTTCTCAAGAGTGATGTGAAGAAATTTAAAGATTTTGAAGAAATGAATCAACCA	439
Db	487	GGTTTTTTTCTTACAGAGTGAAATTTGAAGAAATTTAAGAACTTTAGAAAGAAATGACCTC	546
OY	440	GACATGACAGATGAATTTCTATCAGATTTGGGACATCATGAAG-----	482
Db	547	GACATGACAGATGAATTTCTTTTGGATTTAGGACATCATGAAGGATTTGATGGCAATTA	606
OY	483	-----	482
Db	607	CTGTCTCATTAAGTGAACAATAAAGAAAAATTTGTATTAGCACAACTACTGTGA	666
OY	483	-----GTCTATTAATGACGATCTACTACTACCTAGCTCAAA	517
Db	667	AGAGGAAAAAGATTACCCCATCAAGAGCTCTATTAATGACGAGATCTATTACTCACTCAG	726
OY	518	CAGATGGGGCAGAGTATTGGCGTGAAGAGGAGCCAAAGATCTGACAGAGCTGTGCGAC	577
Db	727	CAGATGGAGCAGAGGTATTGGCGGGAAAAAGAGCCAAAGATCTGACAGAGCTGTGCGAC	786
OY	578	GGAGATTAACATATCTTCAGAAATCCCAAGAGCTGACAGAAAGCCAAAGAGCTAGTGTGA	637
Db	787	GGAGATTAACATATCTTCAGAAATCCCAAGAGCTGACAGAAAGCCAAAGAGCTAGTGTGA	846
OY	638	ATATCAACAAGAGCTGTGGCTATGGCTGACAGCTCCATCATGATGATGTAGTACGCTTATGA	697
Db	847	ATATCAACAAGAGCTGTGGCTATGGCTGACAGCTCCATCATGATGATGTAGTACGCTTATGA	906
OY	698	TTTGCAATATGACCCAGAGGAACACTGCGCTTGGGAATTCACAAATTTGGCGCTACGCTAG	757
Db	907	TTTGCAATATGACCCAGAGGAACACTGCTTGGGAATTCACAAATTTGGCGCTATGCTATAG	966
OY	758	GGGAGTGGGAAACTGTGTTTAAAGCTGTAAAGTGAAGCTGACAGACAGATCTGGCAGCT	817
Db	967	GTGAGTGGGAGACTGTATTAGGCTGTAAAGTGAAGCTGACAGACAGATCTGGCAGCT	1026
OY	818	CCACTGGGCACTTGTGCTAGCTGAAGTAAAGGACAAATAATTTTCAGAGTGTAGAGCTCCCA	877
Db	1027	CCACTGGGCACTTGTGCTAGCTGAAGTAAAGGACAAATAATTTTCAGAGTGTAGAGCTCCCA	1086

QY	878	TTGTGACAGAGTTCATCTCGCTCCATATATTTACCCCTGGCTCCACAGAACCTTGG	937
Db	1087	TTGTAGACAGTCTTCATCCCGCTCCATATTTACCTTGGCTGTACACAGAACCTCG	1146
QY	938	CAGATGCATCTGTACAGTCCATGGTGTATCCTGCAGTGTGGGTATCCAGTTTGTCA	997
Db	1147	CAGATGCATCTGTACAGTGCATGGTGTACCCCTGCAGTGTGGGTGTCTCAGTTTGTCA	1206
QY	998	AGTACTTGATTTCCGCCACAAACCCCTGGCGAAAGGAATAGAAAGGCCCAACGAAGC	1057
Db	1207	AATACTTGATTCGCCGCCAAGCCCTGGCTAGAAAAGAAATAGAAAGCCCAACGAAGC	1266
QY	1058	TAGCGTTGAACATCCAGTATATGGAGTCCATGTTAGACGACAGACAAAGTGGAGCGG	1117
Db	1267	TTGGCTTCAACATCCAGTATATGGAGTCCATGTCAGACGACAGACAAAGTGGGAACAG	1326
QY	1118	AAGCAGCCTTCATCCATTTGAGAAATACACGGTGCACGTTGAGAGACGTTTCACTTC	1177
Db	1327	AAGCTGCCTTCATCCATTTGAGAAAGTGCATGGTGCATGTTGAAAGACATTTTCACTTC	1386
QY	1178	TTGGCTCGAGAAATGCAGTGGATTAAGAAAGGGGTATTTGGCCACAGATGCCCTGCTT	1237
Db	1387	TTGCACGCAATGCAAGTGCACAAAGAAAGGTATTTGGCCACAGATGCCCTTCTTT	1446
QY	1238	TGTTAAAGAGGCAAAAACAAAGTACCCCACTTATGAATTTATAGTATAACTCTATCT	1297
Db	1447	TATTTAAAGAGGCAAAAACAAAGTACCCCAATTTATGAATTTATAGTATAACTCTATTT	1506
QY	1298	CTTGGTCAGCTGCATACATATATGCATATACAGAAATTCACCTCGGGGTGCATCCTGG	1357
Db	1507	CCTGGTCAGCTGCATGCACATGCATGCATACAGAAATTCACCTGTGGAGTGCCTGG	1566
QY	1358	ATATACACTTTCTCTCCAGGAGAGACTTCTAGTGTGTACTTTTTCATCGCAGGTCTGTA	1417
Db	1567	ATATACACTTTCTCTCTCAGGAGAGACTTCTAGTGTGTACTTTTTCATCCAGGTCTGTC	1626
QY	1418	GAGTGTCTTATGAATTCATGCAGAGCGCTGCATCCTGCATGCGCTCTGCGAATTCGTTCTT	1477
Db	1627	GAGTGTCTTATGAATTTATGCAGAACATCAATCCTGAATGCGCTCTGCAAACTTCCATTTCTT	1686
QY	1478	TGGATGACATCTACTATTTTGGAGGCCCAAAATGCCCACACACCAATTCGCAATTTCTCTC	1537
Db	1687	TAGATGACATCTACTATTTTGGGGGCCAGAAATGCCACACATCAATTCGCAATTTATGCTC	1746
QY	1538	ACCAACCTGGAATGAGAGAGAAATCCCATGGAACCTGAGATATTTATGGTGTGGCTG	1597
Db	1747	ACCAACCCCGAACTGCACATGAATTTCCCATGGAACCTGAGATATCTATGGTGTGGCTG	1806
QY	1598	GAAATCACCTGGAGTGGCTATCCTAAAGGTGTTAAACAGAAACTGGGAAGGACGGGCTAT	1657
Db	1807	GAAATCACCTGGAGTGGCTATTTCTAAAGGTGTCAACAGAAATTTGGGAAGGACGGGCTAT	1866
QY	1658	ATCCCTCTACAAAGTTGAGAGAGATAGAAACAGTCAAGTACCCCAATATCCCGAGG	1717
Db	1867	ATCCCTCTACAAAGTTGAGAGAGATAGAAACAGTCAAGTACCCCAATATCTCTGAGG	1926
QY	1718	CTGACAAGTAA 1728	
Db	1927	CTGAGAATAAA 1937	

[illegible]

```
REFERENCE 1 (bases 1 to 1850)
AUTHORS Anderson,B., Wentland,M.A., Ricefrente,J.Y., Liu,W. and Gibbs,R.A.
TITLE A 'double adaptor' method for improved shotgun library construction
JOURNAL Anal. Biochem. 236 (1), 107-113 (1996)
MEDLINE 96207227
PUBMED 8619474
REFERENCE 2 (bases 1 to 1850)
AUTHORS Yu,W., Anderson,B., Worley,K.C., Muzny,D.M., Ding,Y., Liu,W.,
Ricefrente,J.Y., Wentland,M.A., Lennon,G. and Gibbs,R.A.
TITLE Large-scale concatenation cDNA sequencing
JOURNAL Genome Res. 7 (4), 353-358 (1997)
MEDLINE 97264341
PUBMED 9110174
REFERENCE 3 (bases 1 to 1850)
AUTHORS Yu,W., Sargison,J. and Gibbs,R.A.
TITLE Direct Submision
JOURNAL Submitted (05-MAR-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA
FEATURES
source
1. 1850
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="T.M.A.G.E. Consortium clone ID 23830"
/sex="female"
/clisue_type="brain"
/clone_lib="INIB"
/dev_stage="infant"
/note="similar to Homo sapiens mRNA for
N-acetyl-beta-D-glucosaminide encoded by GenBank Accession
Number DB9289"
BASE COUNT 576 a 378 c 374 g 522 t
ORIGIN
Query Match 55.0%; Score 950; DB 9; Length 1850;
Best Local Similarity 91.8%; Pred. No. 4.1e-249;
Matches 1004; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 635 GTATATTCACAAAGGCTGGCTATGGCTGCTCAGCTCCATCATGTAGTCTACTGCTTGA 694
Db 1 GTATATTCACAAAGGCTGGCTATGGCTGCTCAGCTCCATCATGTAGTCTACTGCTTGA 60
QY 695 TGATTCATATGGCCACCACGAACTCCCTTGAATCTCAATTTGGCGCTACGCTA 754
Db 61 TGATTCATATGGCCACCACGAACTCACTCATTTGGAAATTCAGAAATTTGGCGCTATGCTA 120
QY 755 CTGGGGGATGGAACTGTCTTTAGACCTGTAAGTGAGACGTGCACAGACATCTGGCA 814
Db 121 CTGGTGGATGGGAGACTGTATTAGGCTCTTAAGTGAGACATGCACAGACATCTGGCA 180
QY 815 GCTCCACTGGACATTTGGTCAAGTGAAGTAAAGCAAAATGTTCAAGGTGGTGAAGCTCC 874
Db 181 TCTCCACTGGACATTTGGTCAAGTGAAGTGAAGCAAAATGTTCAAGGTGGTGAAGCTTC 240
QY 875 CCATTGTAGACAGTGTTCATCTGCTCCATATTACCCCTGGCTGTGCCAGAAAGCC 934
Db 241 CCATTGTAGACAGTGTTCATCTGCTCCATATTACCCCTGGCTGTGCCAGAAAGCC 300
QY 935 TTCCAGATGCACTTTGTAAGAGTGCATGCTGTCAGTGTGGTGGTATCCAGTTTG 994
Db 301 TTCCAGATGCACTTTGTAAGAGTGCATGCTGTCAGTGTGGTGGTGTCTCAGTTTG 360
QY 995 TCAAGTACTGATTTCGCCCAACACCTGGCTGGAAGAAATGAAAGAGCCACCAAGA 1054
Db 361 TCAAAATACTGATTTCGCCCAACACCTGGCTGGAAGAAATGAAAGAGCCACCAAGA 420
QY 1055 AGCTAGGCTTCAAAACATCCAGTTATTGAGAGTCCATGTTAGACGCACAGACAAGTGGAG 1114
Db 421 AGCTTGGCTTCAAAACATCCAGTTATTGAGAGTCCATGTTAGACGCACAGACAAGTGGAA 480
QY 1115 CGGAGACAGCTTTCATCCATTGAGGAATACAGGTGACGTTGAAGAAAGACTTTCAGC 1174
Db 481 CAGAGAGCTGCTTCATCCATTGAGGAATACAGGTGATGTTGAAGAAAGACTTTCAGC 540
```

```
QY 1175 TTCTGCTCGCAGAAATGCAAGTGGATTAAGAAAGGCTATTTTGGCCACAGATGACCCTG 1234
Db 541 TTCTTGCACCCAGAAATGCAAGTGGATTAAGAAAGGCTATTTTGGCCACAGATGACCCTG 600
QY 1235 CTTTGTAAAGAGGCAAAAGCAAGATGACCCAGTATGATTTATTTAGTATTAACCTA 1294
Db 601 CTTTGTAAAGAGGCAAAAGCAAGATGACCCAGTATGATTTATTTAGTATTAACCTA 660
QY 1295 TCTCTGGTACGCTGACCTACATTAATGATATACAGAAATTTACTTGGGGGTGATCC 1354
Db 661 TTTCTCTGGTACGCTGACCTACATTAATGATATACAGAAATTTACTTGGGGGTGATCC 720
QY 1355 TGGATATACACTTCTCTCCAGGACAGCTTCCAGTGTGATCTTTTCATGCGAGCTT 1414
Db 721 TGGATATACACTTCTCTCCAGGACAGCTTCCAGTGTGATCTTTTCATGCGAGCTT 780
QY 1415 GTAGAGTTGCTTATGAAATCATGCAAGCGCTGCATCTGATGCTTGCAGAACTTCCGT 1474
Db 781 GTGAGTTGCTTATGAAATCATGCAAACTACATCTGATGCTTGCAGAACTTCCAT 840
QY 1475 CTTTGGATGACATCTACTATTTTGGAGGCCCCAAATGCCCAACCAATTTCCATTTATC 1534
Db 841 CTTTGGATGACATCTACTATTTTGGAGGCCCCAAATGCCCAACCAATTTCCATTTATC 900
QY 1535 CTGACCAACCTCGAAGTGAAGGAAATGCCCATGGAACCTGGAGATATTATTTGGTGG 1594
Db 901 CTGACCAACCTCGAAGTGAAGGAAATGCCCATGGAACCTGGAGATATTATTTGGTGG 960
QY 1595 CTGGAATTCATGCGATGCTATCTTAAGGTGTTAAAGAACTGGGAAGACGGGCC 1654
Db 961 CTGGAATTCATGCGATGCTATCTTAAGGTGTTAAAGGTGTTAAAGAACTGGGAAGACGGGCC 1020
QY 1655 TATATCCCTCTCAAAAGTTGAGAGAGATAGAAACAGTCAAGTACCCACATATCCG 1714
Db 1021 TATATCCCTCTCAAAAGTTGAGAGAGATAGAAACAGTCAAGTACCCACATATCTG 1080
QY 1715 AGGCTGACAGTAA 1728
Db 1081 AGGCTGACAGTAA 1094
```

Search completed: November 1, 2002, 22:31:34
Job time : 3121 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:01 : Search time 31 Seconds
(without alignments)
3208.777 Million cell updates/sec

Title: US-09-839-136-2
Perfect score: 3096
Sequence: 1 MRPWGSMRWIMILIFAMGT.....YKVKRIETVKYPTYPEADK 575

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3088	99.7	575	6	P79282 sus scrofa
2	2949	95.3	575	4	O00235
3	2945	95.1	575	6	O9N0M2
4	2903	93.8	575	11	O921U1
5	2892	93.4	575	11	O9WTS2
6	1431.5	46.2	619	5	O9YVY5
7	1420	45.9	276	4	O9P2U6
8	1109	35.6	246	4	O9BYC6
9	948	30.6	818	5	O16882
10	529	17.1	105	4	O9BYC5
11	261	8.4	50	4	O9P2U5
12	140	4.5	1025	10	O9SAAS
13	123.5	4.0	1057	16	O9N0M5
14	123	4.0	1871	10	O9SRH5
15	122	3.9	1200	11	O921B9
16	121.5	3.9	1154	16	O9ZK97

17	119	3.8	1455	5	O9V587
18	117.5	3.8	725	4	O15624
19	117.5	3.8	2363	4	O99968
20	117	3.8	1958	5	O96062
21	116	3.7	544	5	O61196
22	116	3.7	1146	11	O9WVE1
23	116	3.7	1217	11	O9WVE9
24	115	3.7	584	17	O96Y06
25	113.5	3.7	1097	5	O9VIE7
26	113	3.6	1391	11	O922J3
27	112.5	3.6	647	5	O27031
28	112	3.6	978	16	O99735
29	112	3.6	1527	11	O9ES67
30	111	3.6	398	10	O49583
31	111	3.6	570	17	O97U11
32	111	3.6	1011	5	O61639
33	111	3.6	1094	5	O61618
34	111	3.6	1549	5	O20060
35	111	3.6	1680	5	O9U8G1
36	111	3.6	1940	5	O02456
37	110.5	3.6	794	4	O96CM4
38	110.5	3.6	909	1	O57703
39	109.5	3.5	975	13	O987Q5
40	109.5	3.5	1179	17	O59462
41	109.5	3.5	1397	10	O9LP90
42	109	3.5	217	4	O961T4
43	109	3.5	417	17	O9V104
44	109	3.5	822	17	O27154
45	109	3.5	5362	2	O93155

ALIGNMENTS

RESULT 1	ALIGNMENTS
ID P79282	PRELIMINARY: PRT: 575 AA.
AC P79282:	
DT 01-JUN-1998 (TREMBLrel. 06, Created)	
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)	
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
DE ALPHA-(1,6)-FUCOSYLTRANSFERASE (EC 2.4.1.68) (GDP-FUCOSE--GLYCOPROTEIN DE FUCOSYLTRANSFERASE) (FUCOSYLTRANSFERASE 9).	
OS Sus scrofa (pig).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.	
OX NCBI_TaxID=9823;	
RM [1]	
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.	
RP TISSUE=BRAIN:	
RX MEDLINE=97066976; PubMed=8910378;	
RA Uozumi N., Yanagidani S., Miyoshi E., Ihara Y., Sakuma T., Gao C.-X.,	
RA Teshima T., Fujii S., Shiba T., Taniguchi N.;	
RT "Purification and cDNA cloning of porcine brain GDP-Fuc:N-acetyl-	
RT beta-D-glucosaminide alpha1->6fucosyltransferase.";	
RL J. Biol. Chem. 271:27810-27817(1996)	
CC - FUNCTION: TRANSFERS FUCOSE FROM GDP-FUC TO N-LINKED TYPE COMPLEX GLYCOPOLYMERIDES. THE PH OPTIMUM IS 7.	
CC - CATALYTIC ACTIVITY: GDP-L-FUCOSE + N4-(N-ACETYL-BETA-D-GLUCOSAMINYL)-1,2-ALPHA-D-MANNOSYL-1,3-(R(1))-ALPHA-1,6)-BETA-D-MANNOSYL-BETA-N-ACETYL-1,4-D-GLUCOSAMINYL-1,4-N-ACETYL-D-GLUCOSAMINYL)ASPARAGINE - GDP + N4-(N-ACETYL-BETA-D-GLUCOSAMINYL)-1,2-ALPHA-D-MANNOSYL-1,3-(R(1))-ALPHA-1,6)-BETA-D-MANNOSYL-1,4-BETA-N-ACETYL-D-GLUCOSAMINYL-1,4-(ALPHA-L-FUCOSYL-1,6)-N-ACETYL-D-GLUCOSAMINYL)ASPARAGINE.	
CC - PATHWAY: GLYCOSYLATION.	
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND FORM IN TRANS CISTERNAE OF GOLGI.	
CC - TISSUE SPECIFICITY: HIGHEST EXPRESSION IN BRAIN.	
CC EMBL: D86723; BAA13157.1; -;	
DR InterPro: IPR001452; SH3.	
DR Pfam: PF00018; SH3; 1.	
DR SMART: SM00326; SH3; 1.	

DR PROSITE; P550002; SH3; 1.
 KW Transferrase; Glycosyltransferase; Transmembrane; Signal-anchor;
 KM Golgi stack.
 FT DOMAIN 1 9 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 10 25 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 26 575 LUMENAL, CATALYTIC (POTENTIAL).
 SQ SEQUENCE 575 AA; 66229 MW; 0F19D0BC2018F7B CRC64;

Query Match 99.7%; Score 3088; DB 6; Length 575;
 Best local similarity 99.8%; Pred. No. 4e-225;
 Matches 574; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRPTGSMRWIMLILFAWGTLFLYIGGHLVRDNDHSDHSSRELSTLAKLERLKOQNEEL 60
 DB 1 MRPTGSMRWIMLILFAWGTLFLYIGGHLVRDNDHSDHSSRELSTLAKLERLKOQNEEL 60
 QY 61 RRMAGSLRIPEGPIIDGPRAGRVRALEEDPMKAKEDQIENYKKTGKPGKDHILRRRIE 120
 DB 61 RRMAGSLRIPEGPIIDGPRAGRVRALEEDPMKAKEDQIENYKKTGKPGKDHILRRRIE 120
 QY 121 NGAKELMFLQSELEKRLKLNLEGNELORHNADEFSLDGLNHERSIIMTDLYYLSQTDGAGDMR 180
 DB 121 NGAKELMFLQSELEKRLKLNLEGNELORHNADEFSLDGLNHERSIIMTDLYYLSQTDGAGDMR 180
 QY 181 EKEAKDLTELVRRTITVLRNPKDCSKAKKLVCNINKGCGYCOLHNHYVCFMAYGTQRT 240
 DB 181 EKEAKDLTELVRRTITVLRNPKDCSKAKKLVCNINKGCGYCOLHNHYVCFMAYGTQRT 240
 QY 241 LAESHNMWRATGMEVFRPVSECTDRSGSGTGHMGSVEKKNQVVELPIVDSVHPR 300
 DB 241 LAESHNMWRATGMEVFRPVSECTDRSGSGTGHMGSVEKKNQVVELPIVDSVHPR 300
 QY 301 PPTPLAIVEDLADRLVRVHGDPVAVMVVSQFVYLLIRPQWLEKEIEATKLGFKHPRV 360
 DB 301 PPTPLAIVEDLADRLVRVHGDPVAVMVVSQFVYLLIRPQWLEKEIEATKLGFKHPRV 360
 QY 361 GVHVRRDVKGAFAAHPRIEYTVHVEEDFOLLARRMOVDKRVYLATDDPALLKKAATK 420
 DB 361 GVHVRRDVKGAFAAHPRIEYTVHVEEDFOLLARRMOVDKRVYLATDDPALLKKAATK 420
 QY 421 YPTEFTSDNISMSAGLNHRRTENSLRGVILDIHFLSQADFLVCFSSOVCAVAEIMQ 480
 DB 421 YPTEFTSDNISMSAGLNHRRTENSLRGVILDIHFLSQADFLVCFSSOVCAVAEIMQ 480
 QY 481 ALHPDASANFRSLDDIYFEGGPAHNOIAIYHPQRTGEGIPMEPDGIIGVAGNHMDGYP 540
 DB 481 ALHPDASANFRSLDDIYFEGGPAHNOIAIYHPQRTGEGIPMEPDGIIGVAGNHMDGYP 540
 QY 541 KGVNRKLGRTGLPSYKVRKEITVYKPYTPPEADK 575
 DB 541 KGVNRKLGRTGLPSYKVRKEITVYKPYTPPEADK 575

RESULT 2

000235 PRELIMINARY: PRT: 575 AA.

AC 000235;
 DT 01-JUL-1997 (Tremblrel. 04, Created)
 DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DB N-ACETYL-BETA-D-GLUCOSAMINIDE (ALPHA1-6FUCOSYLTRANSFERASE).
 GN ALPHA1-6FUCOT OR F0T8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97279058; Pubmed=9133635;
 RA Yanaguchi N., Uozumi N., Ihara Y., Miyoshi E., Yamaguchi N.,
 RA Taniguchi N.,
 RT "Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-
 glucosaminide:alpha1-6 fucosyltransferase (alpha1-6 fucT) from human

RT gastric cancer MKN45 cells.";
 RL J. Biochem. 121:626-632(1997).
 RN 12)
 RP SEQUENCE OF 278-419 FROM N.A.
 RA Yanaguchi Y.,
 RL Submitted (Dec-1997) to the EMBL/GenBank/DBJ databases.
 RN 13)
 RP SEQUENCE FROM N.A.

RC TISSUE-WHOLE EMBRYO;
 RA Cailleau A., Balanzino L., Candelier J.J., Oriol R., Mollicone R.,
 RT "Differential splice variants of human F0T8 embryonic cDNA."
 RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: D89289; BAA19764.1; -;
 DR EMBL: AF038281; AAB92372.2; -;
 DR EMBL: AF038280; AAB92372.2; JOINED.
 DR EMBL: Y17979; CAA76988.1; -;
 DR EMBL: Y17976; CAA76985.1; -;
 DR EMBL: Y17977; CAA76986.1; -;
 DR EMBL: Y17978; CAA76987.1; -;
 DR InterPro: IP001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR SMART: SM00326; SH3; 1.
 KW Transferrase; Glycosyltransferase.

Query Match 95.3%; Score 2949; DB 4; Length 575;
 Best local similarity 95.5%; Pred. No. 1.3e-214;
 Matches 549; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY 1 MRPTGSMRWIMLILFAWGTLFLYIGGHLVRDNDHSDHSSRELSTLAKLERLKOQNEEL 60
 DB 1 MRPTGSMRWIMLILFAWGTLFLYIGGHLVRDNDHSDHSSRELSTLAKLERLKOQNEEL 60
 QY 61 RRMAGSLRIPEGPIIDGPRAGRVRALEEDPMKAKEDQIENYKKTGKPGKDHILRRRIE 120
 DB 61 RRMAGSLRIPEGPIIDGPRAGRVRALEEDPMKAKEDQIENYKKTGKPGKDHILRRRIE 120
 QY 121 NGAKELMFLQSELEKRLKLNLEGNELORHNADEFSLDGLNHERSIIMTDLYYLSQTDGAGDMR 180
 DB 121 NGAKELMFLQSELEKRLKLNLEGNELORHNADEFSLDGLNHERSIIMTDLYYLSQTDGAGDMR 180
 QY 181 EKEAKDLTELVRRTITVLRNPKDCSKAKKLVCNINKGCGYCOLHNHYVCFMAYGTQRT 240
 DB 181 EKEAKDLTELVRRTITVLRNPKDCSKAKKLVCNINKGCGYCOLHNHYVCFMAYGTQRT 240
 QY 241 LAESHNMWRATGMEVFRPVSECTDRSGSGTGHMGSVEKKNQVVELPIVDSVHPR 300
 DB 241 LAESHNMWRATGMEVFRPVSECTDRSGSGTGHMGSVEKKNQVVELPIVDSVHPR 300
 QY 301 PPTPLAIVEDLADRLVRVHGDPVAVMVVSQFVYLLIRPQWLEKEIEATKLGFKHPRV 360
 DB 301 PPTPLAIVEDLADRLVRVHGDPVAVMVVSQFVYLLIRPQWLEKEIEATKLGFKHPRV 360
 QY 361 GVHVRRDVKGAFAAHPRIEYTVHVEEDFOLLARRMOVDKRVYLATDDPALLKKAATK 420
 DB 361 GVHVRRDVKGAFAAHPRIEYTVHVEEDFOLLARRMOVDKRVYLATDDPALLKKAATK 420
 QY 421 YPTEFTSDNISMSAGLNHRRTENSLRGVILDIHFLSQADFLVCFSSOVCAVAEIMQ 480
 DB 421 YPTEFTSDNISMSAGLNHRRTENSLRGVILDIHFLSQADFLVCFSSOVCAVAEIMQ 480
 QY 481 ALHPDASANFRSLDDIYFEGGPAHNOIAIYHPQRTGEGIPMEPDGIIGVAGNHMDGYP 540
 DB 481 ALHPDASANFRSLDDIYFEGGPAHNOIAIYHPQRTGEGIPMEPDGIIGVAGNHMDGYP 540
 QY 541 KGVNRKLGRTGLPSYKVRKEITVYKPYTPPEADK 575
 DB 541 KGVNRKLGRTGLPSYKVRKEITVYKPYTPPEADK 575

RESULT 3

09NOW2 PRELIMINARY: PRT: 575 AA.

AC Q9N0W2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
GN 6-ALPHA-L-FUCOSYLTRANSFERASE.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Jevaud C., Dupuy F., Maftah A., Michalski J.C., Oriol R., Petit J.M.,
Julien R.;
RT "Ancestral exonic organization of FUT8, the gene encoding the alpha6-
fucosyltransferase reveals successive peptide domains which suggest a
RT particular 3D core-structure for the alpha6-fucosyltransferase
RT family."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF247186; AAF65460.1; -.
DR InterPro; IPR001452; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KM Transferrase: Glycosyltransferase.
SQ SEQUENCE 575 AA; 66492 MW; FE04C6688A5BF540 CRC64;

Query Match 95.1%; Score 2945; DB 6; Length 575;
Best Local Similarity 95.1%; Pred. No. 2.6e-214;
Matches 547; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

QY 1 MRPWGSRWIMLILFANGTLIFYGHLVNRNDHSDHSSRELSTILAKLERLKOONEDL 60
DB 1 MRPWGSRWIMLILFANGTLIFYGHLVNRNDHSDHSSRELSTILAKLERLKOONEDL 60
QY 61 RRMAGSLRIPGPIIDOGSPASGRVRLAEQFMKAKQIENYKQKOTNGCPKDHIELRRRIE 120
DB 61 RRMAGSLRIPGPIIDOGSPASGRVRLAEQFMKAKQIENYKQKOTNGCPKDHIELRRRIE 120
QY 121 NGAKELWFFLOSELKLLKLNLEGNELQRNADEFLSDLGHNSRIMTDLYYLSOTDAGDMR 180
DB 121 NGAKELWFFLOSELKLLKLNLEGNELQRNADEFLSDLGHNSRIMTDLYYLSOTDAGDMR 180
QY 181 EKEADLDELVQRRITTYLQNPDKSKAKKLVNCNKGCGYGCGLHNVVCFPIAIGTQRT 240
DB 181 EKEADLDELVQRRITTYLQNPDKSKAKKLVNCNKGCGYGCGLHNVVCFPIAIGTQRT 240
QY 241 LALESHNRATAGMETVFRPVSETCTDRSGSSSTGMSGEVDKINVQVELPIVDSVHPR 300
DB 241 LALESHNRATAGMETVFRPVSETCTDRSGSSSTGMSGEVDKINVQVELPIVDSVHPR 300
QY 301 PPLYLAVPEDLADRLVNHGDPVAVWWSQFVKYLIRQPMLEKEIEEATKKLGFKHPI 360
DB 301 PPLYLAVPEDLADRLVNHGDPVAVWWSQFVKYLIRQPMLEKEIEEATKKLGFKHPI 360
QY 361 GVHVARTKVGAEAFHPIEEYTVVVEEDFOLLARMQVKKRVYLAIDDDPALLEAKTK 420
DB 361 GVHVARTKVGAEAFHPIEEYTVVVEEDFOLLARMQVKKRVYLAIDDDPALLEAKTK 420
QY 421 YPSYEFISDNTSMAGLHNRRTENSLRGVILDIHFLSQADFLVCTFSSQCVRAVEIEMQ 480
DB 421 YPSYEFISDNTSMAGLHNRRTENSLRGVILDIHFLSQADFLVCTFSSQCVRAVEIEMQ 480
QY 481 ALHPDASANFSLDIIYFGCPNANHQAIIYPHQRTGELIPMEBDIIGVAGNHMDGYP 540
DB 481 ALHPDASANFSLDIIYFGCPNANHQAIIYPHQRTGELIPMEBDIIGVAGNHMDGYP 540
QY 541 KGVNKKLGRGTGLPSYKVKREKETVKKPYTPPEADK 575
DB 541 KGVNKKLGRGTGLPSYKVKREKETVKKPYTPPEADK 575

RESULT 4
Q921U1
ID Q921U1 PRELIMINARY; PRT: 575 AA.
AC Q921U1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN FUCOSYLTRANSFERASE 8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010666; AAH10666.1; -.
KM Transferrase.
SQ SEQUENCE 575 AA; 66556 MW; 226092A8959B3EB7 CRC64;

Query Match 93.8%; Score 2903; DB 11; Length 575;
Best Local Similarity 93.4%; Pred. No. 3.9e-211;
Matches 537; Conservative 14; Mismatches 24; Indels 0; Gaps 0;

QY 1 MRPWGSRWIMLILFANGTLIFYGHLVNRNDHSDHSSRELSTILAKLERLKOONEDL 60
DB 1 MRPWGSRWIMLILFANGTLIFYGHLVNRNDHSDHSSRELSTILAKLERLKOONEDL 60
QY 61 RRMAGSLRIPGPIIDOGSPASGRVRLAEQFMKAKQIENYKQKOTNGCPKDHIELRRRIE 120
DB 61 RRMAGSLRIPGPIIDOGSPASGRVRLAEQFMKAKQIENYKQKOTNGCPKDHIELRRRIE 120
QY 121 NGAKELWFFLOSELKLLKLNLEGNELQRNADEFLSDLGHNSRIMTDLYYLSOTDAGDMR 180
DB 121 NGAKELWFFLOSELKLLKLNLEGNELQRNADEFLSDLGHNSRIMTDLYYLSOTDAGDMR 180
QY 181 EKEADLDELVQRRITTYLQNPDKSKAKKLVNCNKGCGYGCGLHNVVCFPIAIGTQRT 240
DB 181 EKEADLDELVQRRITTYLQNPDKSKAKKLVNCNKGCGYGCGLHNVVCFPIAIGTQRT 240
QY 241 LALESHNRATAGMETVFRPVSETCTDRSGSSSTGMSGEVDKINVQVELPIVDSVHPR 300
DB 241 LALESHNRATAGMETVFRPVSETCTDRSGSSSTGMSGEVDKINVQVELPIVDSVHPR 300
QY 301 PPLYLAVPEDLADRLVNHGDPVAVWWSQFVKYLIRQPMLEKEIEEATKKLGFKHPI 360
DB 301 PPLYLAVPEDLADRLVNHGDPVAVWWSQFVKYLIRQPMLEKEIEEATKKLGFKHPI 360
QY 361 GVHVARTKVGAEAFHPIEEYTVVVEEDFOLLARMQVKKRVYLAIDDDPALLEAKTK 420
DB 361 GVHVARTKVGAEAFHPIEEYTVVVEEDFOLLARMQVKKRVYLAIDDDPALLEAKTK 420
QY 421 YPSYEFISDNTSMAGLHNRRTENSLRGVILDIHFLSQADFLVCTFSSQCVRAVEIEMQ 480
DB 421 YPSYEFISDNTSMAGLHNRRTENSLRGVILDIHFLSQADFLVCTFSSQCVRAVEIEMQ 480
QY 481 ALHPDASANFSLDIIYFGCPNANHQAIIYPHQRTGELIPMEBDIIGVAGNHMDGYP 540
DB 481 ALHPDASANFSLDIIYFGCPNANHQAIIYPHQRTGELIPMEBDIIGVAGNHMDGYP 540
QY 541 KGVNKKLGRGTGLPSYKVKREKETVKKPYTPPEADK 575
DB 541 KGVNKKLGRGTGLPSYKVKREKETVKKPYTPPEADK 575

RESULT 5
Q9WTS2
ID Q9WTS2 PRELIMINARY; PRT: 575 AA.
AC Q9WTS2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
GN ALPHA-L-6-FUCOSYLTRANSFERASE.


```
OY 61 RMAGSLRIPGPIIDGPGASGRVAL-----EEOFMKAKEQIENYKK 102
DB 74 KQIDELMSDO--LDKOSAMKLVQRLENDALNPKLAPVAGPEPESMFESAPADLRGMN 131
OY 103 QTKNGPGK-----DHEILRRRIENGAKELMFLOSELKRLKML-----EG 142
DB 132 VAEGAPNPLEAGVPDGHGFEPSELEFTRRRRIQOTIGELIMNFSESLCKVRKAVAGHAS 191
OY 143 NELORHADEFSLDGHHERSIMTDLIYLSQTDGAGDMREKAKDITELVQRITTYLQNP 202
DB 192 ADLESISINQVLQGAENRRLSLDMEERKOSDGEAMRHKKEARDISDVQRRLHLQDPS 251
OY 203 DCSKAKKLVQCNINKGCGCOLHNHYVCFMTAYGQPTLALSHNMWRATGEMETVPRV 262
DB 252 DCONARKLVCKLKNKCGGCOLHNHYVCFIYAVATERLLKSKGRMYHKGGEVEFQPV 311
OY 263 SETCTDRSGSSGTGHWGSEKKDNQOVVELPIVDSVHRPRPYLPLAVPEDLADRLVRHGD 322
DB 312 SNSCHDACTANTMYNMPK---PNTQVILVPLIIDSIMPRPYLPLAVPEDLAPRLKRLHGD 368
OY 323 PAVMNVSGPVKYLIRPQWLEKEIEATKKGFKHPIVIGVHVRTPDKVGAEAHFPIEBY 382
DB 369 PIVMNVGQFLAYLLRPOPTTRDFLTSGMRNLGWERPIVGVHVRTPDKVGTAAACHSVEEY 428
OY 383 TVHVEDEFLARMQVQDKKRYLATDDPDLKKEAKTKYPSYEFTSDSISMSAGLHNR 442
DB 429 MTVEEDYRTLELVNGSTVARKLFLASDAQVLEAKRKYPQOITIGDEPVARMAVSSTRY 488
OY 443 TENSRLGYLDLHFLSQADFVLCFSPSOVCRAVEIEMQALHPDASANFRSLDDIYFSGP 502
DB 489 TDTALNGIILDLHLSMDHLCTFSSQVCRAVEIEMQTMVDAHNRKSLDDIYVYGGO 548
OY 503 NAHNOIATYHPQRTGEGIPMEBGDIIGVAGNHMDGTCVKNRKLGKRGYLPSTVYREKI 562
DB 549 NAHNRVYIAHNRPTHELDQLRVGDLVSVAGNHMDGNSKKNRTNQGCLPSPFVEEKV 608
OY 563 ETVKPTY 570
DB 609 DTAKLPLY 616

RESULT 7
O9P2U6
ID O9P2U6 PRELIMINARY: PRT: 276 AA.
AC O9P2U6:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ALPHA1.6 FUCOSYLTRANSFERASE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20275614; PubMed=10814706;
RA Yamaguchi Y., Ikeda Y., Takahashi T., Ihara H., Tanaka T., Sasao C.,
RA Uozumi N., Yanagidani S., Inoue S., Fujii J., Taniguchi N.;
RT "Genomic structure and promoter analysis of the human alpha1,6-
RT fucosyltransferase gene (FUT8).";
RL Glycobiology 10:637-643(2000).
DR EMBL; AB032572; BAA92858.1;
DR EMBL; AB032572; BAA92858.1; JOINED.
DR EMBL; AB032568; BAA92858.1; JOINED.
DR EMBL; AB032569; BAA92858.1; JOINED.
DR EMBL; AB032570; BAA92858.1; JOINED.
DR EMBL; AB032571; BAA92858.1; JOINED.
KW Transferase; Glycosyltransferase.
FT NON_TER 276
FT SEQUENCE 276 AA; 32182 MW; 18BEC2AC59152E61 CRC64;

Query Match 45.9%; Score 1420; DB 4; Length 276;
Best Local Similarity 95.7%; Pred. No. 1.9e-99;
```

```
Matches 264; Conservative 2; Mismatches 10; Indels 0; Gaps 0;
OY 1 MRPTGSRWIMLILFANGTLFTYIGHLVNDNDHSHSSSELKYLAKLERLQONDL 60
DB 1 MRPTGSRWIMLILFANGTLFTYIGHLVNDNDHSHSSSELKYLAKLERLQONDL 60
OY 61 RMAGSLRIPGPIIDGPGASGRVAL-----EEOFMKAKEQIENYKKOTRNGKPGKHETLRRIE 120
DB 61 RMAGSLRIPGPIIDGPGASGRVAL-----EEOFMKAKEQIENYKKOTRNGKPGKHETLRRIE 120
OY 121 NGAKELMFLOSELKRLKNEGNELORNADEFSLDGHHERSIMTDLIYLSQTDGAGDMR 180
DB 121 NGAKELMFLOSELKRLKNEGNELORNADEFSLDGHHERSIMTDLIYLSQTDGAGDMR 180
OY 181 EKEAKDITELVQRITTYLQNPDKCSKAKKLVQCNINKGCGCOLHNHYVCFMIAVGTORT 240
DB 181 EKEAKDITELVQRITTYLQNPDKCSKAKKLVQCNINKGCGCOLHNHYVCFMIAVGTORT 240
OY 241 LALSHNMWRATGEMETVPRVSETCTDRSGSSGTGH 276
DB 241 LALSHNMWRATGEMETVPRVSETCTDRSGSSGTGH 276

RESULT 8
O9BYC6
ID O9BYC6 PRELIMINARY: PRT: 246 AA.
AC O9BYC6:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ALPHA 1.6-FUCOSYLTRANSFERASE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=RETINA;
RA Yamaguchi Y., Ikeda Y., Ookawara T., Suzuki K., Taniguchi N.;
RT "The splicing variant of the human alpha 1,6-fucosyltransferase.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AB049740; BAB40929.1;
KW Transferase; Glycosyltransferase.
FT NON_TER 1
FT SEQUENCE 246 AA; 28310 MW; BA69C8D960458E14 CRC64;

Query Match 35.8%; Score 1109; DB 4; Length 246;
Best Local Similarity 94.9%; Pred. No. 5.2e-76;
Matches 206; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
OY 63 MAGSLRIPGPIIDGPGASGRVAL-----EEOFMKAKEQIENYKKOTRNGKPGKHETLRRIENG 122
DB 1 MAESLRIPGPIIDGPGASGRVAL-----EEOFMKAKEQIENYKKOTRNGKPGKHETLRRIENG 60
OY 123 AKELMFLOSELKRLKNEGNELORNADEFSLDGHHERSIMTDLIYLSQTDGAGDMREK 182
DB 123 AKELMFLOSELKRLKNEGNELORNADEFSLDGHHERSIMTDLIYLSQTDGAGDMREK 120
OY 183 EAKDITELVQRITTYLQNPDKCSKAKKLVQCNINKGCGCOLHNHYVCFMIAVGTORTLA 242
DB 183 EAKDITELVQRITTYLQNPDKCSKAKKLVQCNINKGCGCOLHNHYVCFMIAVGTORTLI 180
OY 243 LSHNMWRATGEMETVPRVSETCTDRSGSSGTGHWSG 279
DB 181 LSHNMWRATGEMETVPRVSETCTDRSGISGTGHWSG 217

RESULT 9
O16882
ID O16882 PRELIMINARY: PRT: 818 AA.
AC O16882:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
```



```
QY 421 YPSYEFISDINSWSAGLHNRTEENSLRGVLLDHFILSQADFLVCTFSSQ 470
DB 1 YPNEEFISDINSWSAGLHNRTEENSLRGVLLDHFILSQADFLVCTFSSQ 50

RESULT 12
Q9SAAS PRELIMINARY; PRT; 1025 AA.
AC Q9SAAS;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE F25C20.13 PROTEIN.
GN F25C20.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Lee J.M., Kremenetskaya I., Luros J., Ngan I., Liu A.,
RA Gonzalez A., Alcafi H., Araujo R., Chao Q., Conn L., Conway A.B.,
RA Dunn P., Hansen N., Hultzer L., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.,
RT "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007296; AAD30251.1;
SQ SEQUENCE 1025 AA: 116411 MW: 08C952A2032BA1E4 CRC64;

Query Match 4.5%, Score 140; DB 10; Length 1025;
Best Local Similarity 17.8%, Pred. NO. 0.072;
Matches 110; Conservative 80; Mismatches 193; Indels 234; Gaps 27;

QY 3 PWTGSRMRIMILIFAMGILFF-YI-----GGLVPRNDSDHSSRELKILAEELK 54
DB 360 PKSGNW-WPAEVLVPGGALVDWVADGPRPGAFLYDNNQYODFHALVPQKIPELVWLE 418
QY 55 QONEDLRMAGSLRIPREGRITDQSPASGRVRLAEQFMKAK-EQIENYKQTFKNGPKDHE 113
DB 419 EENMFRLQED-----RRLKEVMAKMEKTRALKAETKERTLKKFL 461
QY 114 ILRR-----RIENGAKELMFF-----LOSELKKL 137
DB 462 LSQKDVVTEPLEIQAGNPVTVLYNPANTVLYNGKPEVWFGSGFNMTHTRLPQKMEA 521
QY 138 KNLGNELEORNA---DEFLISLGHHERS-----162
DB 522 TDDESSHVKTAKVPLDLYAMDEVFSEKEDGIFDNKKGLDHLRVGSGISKEPRLIYH 581
QY 163 IMTDLYVLSQTDGAGDMEKEAKDLTELVRRIYVLPQKDCSK---AKKLVCN-----213
DB 582 IAVENAPLAKVGGLDVYVTSLSRAVQEL-NHNVDVLPFKYDDIKINFPKDLQFNSTYHMG 640
QY 214 -----INKGCGYGC-----LHHVVYCEMIAYGTQRT 240
DB 641 GTEIKVWGHKEGLSVYELDPQNGLFQRCGVYGCADADGREGFCHALFEFLQCGFHPD 700
QY 241 LALESHNNRYATGWEYVPRVSEFCTDRSSSTGHWGSEVQDKVQ---VVELEPLVDSVH 298
DB 701 I-LHCHDWSAPVSW--LE-----KDHVTOGLIKITRIVFTIH 735
QY 299 -----PRPYL-PLAVPEDLADRLVAVHG-----DPAVW--WWS 329
DB 736 NLEFGANNIGKAMTFADKATTVSPFYTAKEVAGNSVISAHLVYFHLINCLIDPDIDIPND 795
```

```
QY 330 QPVKYLIRPQWLEKEI-----EATKRLGK---HPYIGVHVNRITDKVGAENAFHP 378
DB 796 NFI-----PVPTSENVVEGKRAAKEELQNLGKLSADPFPVGIITRTJHOKGIHLJKHA 850
QY 379 I-----EEYTVHEEDFOLLARRMOV---DKKRVYLATDDPALKEAKTKY 421
DB 851 IWRILERNGOVYLLGASAPDPRIQNDPVALANQHLSSHCDRRRLVLTIDEPL-----SHLIT 906
QY 422 PSEYEFISDINSWSAGL 438
DB 907 AGADFLVPSIFEECGCL 923

RESULT 13
Q99UR5 PRELIMINARY; PRT; 1057 AA.
AC Q99UR5;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CARBAMOYL-PHOSPHATE SYNTHASE LARGE CHAIN.
GN PYRAB OR SA1046 OR SAV1203.
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OC NCBI_Taxid=158879, 153878;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
RX MEDLINE=21311952; Pubmed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K., I., Nagai Y., Iian J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hoshoyama A.,
RA Mizutani-Ui Y., Takehashi N.K., Sawano T., Inoue R.-I., Kato C.,
RA Sekimizu K., Hiraoka H., Kihara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiratazu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RL aureus."
RT Lancet 357:1225-1240(2001).
DR EMBL: AP003132; BAB42298.1;
DR EMBL: AP003361; BAB57365.1;
DR HSSP; P00968; ICS0.
DR InterPro: IPR000901; CPsase.
DR InterPro: IPR000169; Thiolprot_act_site.
DR pfam: PF00289; CPsase_L_chain; 2.
DR pfam: PF02785; CPsase_L_D2; 2.
DR pfam: PF02787; CPsase_L_D3; 1.
DR PRINTS; PR00098; CPsase.
DR PROSITE; PS00866; CPsase.
DR PROSITE; PS00867; CPsase_2; UNKNOWN_2.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1057 AA: 117171 MW: E3E179EF0591F0F8 CRC64;

Query Match 4.0%, Score 123.5; DB 16; Length 1057;
Best Local Similarity 18.9%, Pred. NO. 1.3;
Matches 115; Conservative 85; Mismatches 221; Indels 189; Gaps 25;

QY 43 LSKILAKLERLKQONEDLRNAGS-----LRIFEGPID-----OGP 78
DB 318 IAKLAAKIANGVLLIDEMLNPTGTGYAEPFLDVIKIPRPFDKREKREGELGTQMK 377
QY 79 ASGRVAL-----EEOFMKAKEQIEN--YKQTKNGPGKDHELKRRITENGAKELFFIQS 132
DB 378 ATGEVMAIGRYEESILKAIRISLEYGVNHLGLPNCESFDLPYIKERISHODERLFFIGE 437
QY 133 ELKRLKNLEGNELQIHAEDEF-----LSDLGHERSLMPTDLYVLSQTDGAGDMEKEAK 185
```

Db 438 AIRGTTLEEHNMQIDYFEELKPNIDIEHQLKEHOGDLEYLKYANDYG-FSDKITA 496
186 DLTELVRRTTYLQNPDKDSKAKLVNINKGCGYCOLHHVYVCFMAYGQRTALBS 245
497 HRENMTBEERYQQLRMENDIKPVYKMY----- 522
246 HNRVATGMEVFRPVSETCTDRSGSGTGHMSGEVKKNVQVY---ELPIVDSVHPRP 301
523 -----DTCAEFESSTPYYGYGTETEENESIVTDKEKILYLGSG----- 560
302 PLYPLAVPEDLADRLVRYVNGDRAVMMVSO--FYKYLIRPQ-----WLEKEL 346
561 ---PIRIGGVEEDYATVH--AVMAIQAGYEAITVNNPETVSTPDSISDKLYFEPLT 614
347 EEAATKLGKFNPIGVHVRRTDKVGAEAAPHREETVH-----VEEDF 390
615 EEDVNMNINERKGVVY---QFGGQTAINLADKLAKKGVKLTGSLLENLAEDRKKE 670
391 QLLARMQVDDKKRVYLATD-DPALKEAKTKY---PSTYF-----ISDMSISWAGL 438
671 EALLRKINVPQPGKSATSPFEALANAAEIGYPVVVRPSVYLGRAEMEIVDN---DKEL 726
439 HNNYTE---NSRGVILDIHFLISQADFLVCTPSSQVCSVAVYIMALHPDASANRSID 494
727 ENYMTQAVKASPEHPVLVD-RYLTGKEIEVDALDGETVYIIPGIMHIER----- 775
495 DIYFEGCPNANOIAIYHPQRTGEGE-----IPMEPG-DITGVAGNMW---DG-YP 540
776 ----ACVHSGDSIAVYPPQTLTDEDELATLEDTITKLGLNITGLINQFYAHNGVYV 830
49 541 KGVNRKLGRT 550
Db 831 LEVNPRESSRT 840

RESULT 14
Q9SRD5 PRELIMINARY: PRT: 1871 AA.
ID Q9SRD5: 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PUTATIVE HEAT SHOCK PROTEIN. 53413-59028.
GN F28016.15.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
RA Maitl R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,
RA Barnstead M.E., Bowman C.L., White O., Niernan W.C., Fraser C.M.;
RA Arabidopsis thaliana chromosome 1 BAC F28016 genomic sequence.*;
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC010718; AAF04452.1; -
DR InterPro: IPR002068; CrystalLin_HSP20.
DR Pfam: PR00011; HSP20; 1.
DR PROSITE: PS01031; HSP20; 1.
KW Heat shock.
SQ SEQUENCE 1871 AA; 216899 MW; 52300C21F3DBF5A1 CRC64;

Query Match 4.0%; Score 123; DB 10; Length 1871;
Best Local Similarity 25.0%; Pred. No. 3.3;
Matches 47; Conservative 40; Mismatches 73; Indels 28; Gaps 9;

49 24 YIGCHLVNDNDSDHSSRELK-ILAKLERLQONEDLRMGSLRIIPGPIDQGPASGR 82
1065 HNGG-----EDNNDHKEEOKENVIKAE-LNTEEDSPFKVETEIQDNGELKRSNVQAK 1118
83 VRALEE-----QPKAKKEQIENYKKQTKNG-----PGKDHEI--LRRIENGAKELWFF 129

Db 1119 RQTEEKDKTRAMEKNEVER-RKQTKDGLKLEGEDEPPELGCHERRGEEDRIEEL--- 1174
130 LQSEL-----KTLKNLEGNLQRNADEFLSDLGHNEHSINTDLYLISQTDGACADMEKEAK 185
1175 VETEISDHEKEKKKDEDEYILRSQDTGKVLDGERERRSKORIKHSVEDEIDQDEDAE 1234
49 186 DLTELVR 193
Db 1235 EAAAVYSR 1242

RESULT 15
Q921B9 PRELIMINARY: PRT: 1200 AA.
ID Q921B9: 01-DEC-2001 (TREMblrel. 19, Created)
AC Q921B9: 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE NUCLEAR PORE COMPLEX-ASSOCIATED PROTEIN TPR (FRAGMENT).
GN TPR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=129/OLA;
RA Sandblad L., Hunziker A., Cordes V.C.;
RT "Evolutionarily conserved mouse tpr is a single-copy gene located on
RT chromosome 1.";
RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ298076; CAC40701.1; -
FT NON_TER 1200
SQ SEQUENCE 1200 AA; 138973 MW; 616663FBA6B267D7 CRC64;

Query Match 3.9%; Score 122; DB 11; Length 1200;
Best Local Similarity 21.1%; Pred. No. 2.1;
Matches 101; Conservative 53; Mismatches 199; Indels 126; Gaps 17;

49 23 FYIGCHLVNDNDSDHSSRELKILAKLERLQONEDLRMGSLRIIPGPIDQGPASGR 82
55 FEIEKRLSOSQRLVYETRECONLRLEKLNQVVLTEKTELE-----TAADR 105
83 VRALEEFKAKKEQIENYKKQTKNGPKDHLRRIENGAKELWFFQSELKLT----- 137
106 NIGIQSOFRAHELEAEKRDLL--RTNERLSQVE-----YLIEDVKRLNEKXL 153
49 138 -KNLEGNELQRNADEFLSDLGHNEHSIMTDLYLISQTDGAGDMEKEAKDLTELVRRT 196
154 ESNTTKGELQLKLE-----LQASDVAVYRREKRLREQEKELLNONS 195
197 YLQNPDKDSKAKLVNINKG---CGYGQCLHHVYVCFMAYGQRTALLESNNRYATG 253
196 WNTLELTKTDELLALGKREKNELIELKCNLEN-----KKEEVLRLEQMGKLTG 246
254 GMEVFRPVSETCTDRSGSGTGHMSGEVKKNVQVVELPI-----VDSVHPRPYPL 306
247 N-EHLQKHEDLLTKLKEKKEQQAQSMEEKFNEMLNHIKLSLYKSAADDSKAKSNELTR 305
49 307 AVE-----DLADRLVRYVNGDRAVMMVSOQYKYLIRPQMLEKIEEATYKL 353
306 AVDELHKLKLEAGEANKTIDHLLQVEESK-----DQMEKEMLEKIGKLEKELNANDLL 360
354 GKHNRYIGVHVRTDKVGA-----EAAHP-----IEEYTVHVEEDQL 392
361 S-----ATRKKGAILSEELAAASPTAAAVAKIVKGMKLTLEYLNVYVETQDL 409
49 393 LARRMQVDDKKRVYLATDPPALKEAKTKY-----SYEISDMSISWAGLHNNYTE 444
410 LLEKO--ENKRIKTYLDE--YKVEYAKAPILKROREETERAKAVASLSAKLEQAKME 464

Wed Nov 6 14:29:26 2002

us-09-839-136-2.rspt

Page 9

Search completed: October 31, 2002, 13:09:39
Job time : 38 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:00 : Search time 14 Seconds

(without alignments)
1590.267 Million cell updates/sec

Title: US-09-839-136-2

Perfect score: 3096
Sequence: 1 MRPWGSMRWIMLILFAMGT.....YKVRKIEIVKYPYPEADK 575

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122.5	4.0	1427	1 REST_HUMAN	P30622 homo sapien
2	117.5	3.8	2349	1 TPR_HUMAN	P12270 homo sapien
3	117	3.8	1938	1 MYSD_CAEEL	P02567 caenorhabdi
4	114.5	3.7	845	1 SCPL_MESAU	Q60563 mesocricetu
5	112	3.6	352	1 SH32_MOUSE	Q62420 mus musculu
6	111.5	3.6	978	1 RA50_AQUAE	Q67124 aquilex aeo
7	111.5	3.6	1071	1 CARB_BACSU	P25994 bacillus su
8	110.5	3.6	1074	1 PLD1_HUMAN	Q13393 homo sapien
9	110	3.6	976	1 SCPL_HUMAN	Q15431 homo sapien
10	110	3.6	993	1 SCPL_MOUSE	Q62209 mus musculu
11	110	3.6	3210	1 CENF_HUMAN	P49454 homo sapien
12	109.5	3.5	480	1 STEA_METUA	Q57911 methanococc
13	109	3.5	214	1 OSF1_HUMAN	Q92882 homo sapien
14	109	3.5	324	1 NOD2_BRAJA	Q45271 bradyrhizob
15	108.5	3.5	1014	1 NEBL_HUMAN	Q76041 homo sapien
16	106.5	3.4	900	1 MANB_HUMAN	Q93324 caenorhabdi
17	106	3.4	352	1 SH32_HUMAN	Q99962 homo sapien
18	106	3.4	1969	1 MYSA_CAEEL	P12844 caenorhabdi
19	105.5	3.4	646	1 HS70_THEIN	P16019 theileria a
20	105.5	3.4	1961	1 MYH9_RAT	Q68812 rattus norv
21	105	3.4	673	1 LOX5_HUMAN	P09917 homo sapien
22	105	3.4	880	1 RA50_HUMAN	Q9uzc8 pyrococcus
23	105	3.4	919	1 RA50_AERPE	Q9YI21 aetopyrum p
24	103.5	3.3	1966	1 MYSB_CAEEL	Q05256 caenorhabdi
25	103.5	3.3	1976	1 MYHA_BOVIN	Q27991 bos taurus
26	103	3.3	887	1 VAV5_SCHPO	Q10213 schizosacch
27	103	3.3	1230	1 UGSA_SOLTRU	Q43846 solanum tub
28	102.5	3.3	1960	1 MYH9_HUMAN	P35579 homo sapien
29	102	3.3	215	1 OSF1_MOUSE	Q62422 mus musculu
30	102	3.3	519	1 ARH5_HUMAN	Q11774 homo sapien
31	102	3.3	733	1 MOTB_STRCM	Q05065 streptomyce
32	101.5	3.3	529	1 VSM6_TRYBB	P26334 trypanosoma
33	101.5	3.3	723	1 MY5B_MOUSE	P21271 mus musculu

ALIGNMENTS

RESULT 1	ID	REST_HUMAN	STANDARD:	PRT: 1427 AA.
AC	P30622:			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Restin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-Sternberg intermediate filament associated protein).			
GN	RSN.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Peripheral blood monocytes;			
RX	MEDLINE=92289675; PubMed=1600942;			
RA	Bilbe G., Delabie J., Brueggem J., Richener H., Asselbergs F.A.M.,			
RA	Cerletti N., Sorg C., Odink K., Tarsay L., Miesendanger W.,			
RA	de Wolf-Peters C., Shipman R.,			
RT	"Restin: a novel intermediate filament-associated protein highly			
RT	expressed in the Reed-Sternberg cells of Hodgkin's disease.";			
RL	EMBO J. 11:2103-2113(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92405160; PubMed=1356075;			
RA	Pierre P., Scheel J., Rickard J.E., Kreis T.E.;			
RT	"CLIP-170 links endocytic vesicles to microtubules.";			
RL	Cell 70:887-900(1992).			
CC	- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.			
CC	- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE			
CC	CYTOSKELETON.			
CC	- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A			
CC	SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.			
CC	- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERBERG CELLS			
CC	OF HODGKIN'S DISEASE.			
CC	- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: X64838; CAA6050.1; -			Q03410 rattus norv
DR	EMBL: M97501; AAA35693.1; -			Q15642 homo sapien
DR	PIR: S22695; S22695.			P11532 homo sapien
DR	MTM: 179838; -			O67032 aquilex aeo
DR	InterPro: IPR000938; CAP-GLY.			Q9bxi6 homo sapien
DR	InterPro: IPR001878; ZnF_CCHC.			Q92280 mus musculu
DR	PIR: PF01302; CAP-GLY; 2.			Q9r269 mus musculu
DR	SMART: SMO0343; ZnF_C2HC; 1.			Q9usi6 schizosacch
				P27502 rice tungro
				P48725 mus musculu
				Q28641 oryctolagus
				Q97wn0 sulfolobus

```

DR PROSITE: PS00845; CAP-GLY-1; 2.
DR PROSITE: PS50245; CAP-GLY-2; 2.
KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
FT DOMAIN 78 120 CAP-GLY 1.
FT DOMAIN 143 204 SER-RICH.
FT DOMAIN 232 274 CAP-GLY 2.
FT DOMAIN 304 331 SER-RICH.
FT DOMAIN 350 1342 COILED COIL (POTENTIAL).
FT DOMAIN 1408 1421 CCHC-BOX.
FT VARSPLIC 457 491 MISSING (IN SHORT ISOFORM).
FT CONFLICT 1069 1069 D -> E (IN REF. 2).
SO SEQUENCE 1427 AA; 160989 MW; 0A4F16DD9425448 CRC64;

Query Match
Best Local Similarity 4.0%; Score 122.5; DB 1; Length 1427;
Matches 104; Conservative 18.3%; Pred. No. 1;
Mismatches 100; Indels 169; Gaps 22;

OY 40 SRLSLIYLALEKLEKQONEDLRMGSLRIPGPIDQSPASGRVRLAE----- 88
DB 588 SKENESILSKSLHANKNSNDV-----IALMKSKLETAISHQ-QAMEELKVSFGKLGST 640
OY 89 ---QFMKAKEQIE---NYKKOTKN-----GRGDHEILRRRIENGAKELMFPIQ 131
DB 641 ETAEFEALKTQLEKMLDIOHEIENLQNOODSRAAHAKEMALRAKMLKVIKKEKNSLE 700
OY 132 S-----ELKKLNLEG--NELQHNADFLSLDGHNR 161
DB 701 AISKLDKADQHLVEMEDTLNKLQEAELKVELEVLAQAKCNQGTVIDNFTSQLATEE 760
OY 162 SITDLYLSQTDGADMDREKAKDITELVQRIITYLQNPK--DCKAKKL----- 210
DB 761 KLL-DLDALRKASSEKSEKSKLRQOLEAQAQIKHLETKNAESSKASSITRELOREL 819
OY 211 -VCNINMGCGYGGQHLNHYVVCFMIAVGTOFTALFESHNMRYATGCMETVPRPVSECTDR 269
DB 820 KLTNLDENLSEVSQVKEI-----EKELOILKEKFASEAVORSQOET----- 866
OY 270 SGSSTGHWGSEVKKVQVVELPIVDSVHPRPYRLP-----AVBEDLALVRAHGDP 323
DB 867 -----VNKLHQKEQFNMLSLDEKLRENLDMEAKPREK- 901
OY 324 AVWWSQFVKYLIRPQWLEKEIEEATK-----KLGKRFHYIGVHVRITOK 369
DB 902 -----DEREQDLIKAKEKLENDIAELIMKMSGDSSQLTKMDELRIKREYVELQDLKLR 956
OY 370 VGEAAFA--HPIREYVHVEDFQFLARRNOVKRRVYLATDPPALKEAKTKYPSYEPI 427
DB 957 ANENASFLQKSIDMTYVKAQSQOEAQAKHKEEKLEKRLSD--LEKKMETSHNOCQEL 1014
OY 428 S---DNSISMSAGLHNR-----TENSILR-----GVILDIHFL-SQADEFLVCT 466
DB 1015 KANERATSETKTKHEEILQNLQKTLTDREDKLGAAREKMSGLOLELELRKQAD---KA 1071
OY 467 FSSQVCRVAVETIQALHPDASANFRSLDD 495
DB 1072 KAAQTAEADAMQIMEQMTKETETLASIED 1100

RESULT 2
TPR_HUMAN STANDARD; PRT; 2349 AA.
AC P12270;
DT 01-OCT-1989 (rel. 12, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Nucleoprotein TPR.
GN TPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE-93064711; PubMed-1437155;
RA "Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RL extensive coiled-coil regions and an acidic C-terminal domain.";
RN Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.
RX MEDLINE-95096166; PubMed-7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
RP SEQUENCE OF 1-142 FROM N.A.
RX MEDLINE-88262257; PubMed-3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RT "tpr homologues activate met and raf.";
RL Oncogene 2:617-619(1988).
CC -I- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P62.
CC -I- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
CC BRAIN. LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
CC -I- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
CC OR RAF GENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X66397; CAA47021.1; -
DR EMBL: Y00672; CAA68681.1; -
DR PIR: S00928; S00928.
DR MIM: 189940; -
KW Nuclear protein; Proto-oncogene; Chromosomal translocation;
KW Coiled coil; Proco-oncogene; Chromosomal translocation;
FT DOMAIN 78 360
FT DOMAIN 422 571 COILED COIL (POTENTIAL).
FT DOMAIN 575 628 COILED COIL (POTENTIAL).
FT DOMAIN 758 805 COILED COIL (POTENTIAL).
FT DOMAIN 834 869 COILED COIL (POTENTIAL).
FT DOMAIN 934 979 COILED COIL (POTENTIAL).
FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
FT DOMAIN 1196 1241 COILED COIL (POTENTIAL).
FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
FT DOMAIN 527 530 POLY-SER.
FT DOMAIN 1833 1836 POLY-GLU.
FT DOMAIN 1957 1964 POLY-ASP.
FT DOMAIN 2295 2298 POLY-SER.
SO SEQUENCE 2349 AA; AFDD6885CEDCA9EF CRC64;

Query Match
Best Local Similarity 3.8%; Score 117.5; DB 1; Length 2349;
Matches 96; Conservative 21.3%; Pred. No. 4; Mismatches 184; Indels 119; Gaps 16;

OY 23 FYGGLVNDNDHSDSSRELKILAKLEKQONEDLRMGSLRIPGPIDQSPASGR 82
DB 55 FETELKLSHSGELVNETRECSRLRELEKLNQALKALTEKKKELEI-----AQR 105
OY 83 VRALAEQFMKAKEQIEENYKKOTKNQNGKDHILRRRIENGAKELWFFLQSELUKLT----- 137

```

```

DB 106 NIAIOSOPTRKEELEAKRD-----LARTNERLSQEL-EYLTEDVKRLNEKX 153
OY 138 -KNIEGNELOHNADEBLSDCJHNEHSIMTDLYISQTDGACDWRKEAKDLTELQVRIT 196
DB 154 ESNPTTKGLOQLKDE-----LQASDVSKYREKRLDEQKELHSQNT 195
OY 197 YLQPKDCSKAKKLVCNINKG---CGYGCOLHHVYVCPMAVGTORTLALSHMNRVGTG 253
DB 196 WLNTLTKTKTDELLALGKNGNEILELKCNLEN-----KKEEVSRLQDMNGMLKTS 246
OY 254 GMEVTFRPVSETCTDRSGSTGHSGEYKDKNVQVELPI-----VDSVHPRPYPL 306
DB 247 N-EHLQKIVEDLLKTKAKKQOASMEKPHNELNAHKLKSNLYKSADSEAKSNELTR 305
OY 307 AVPE-----DLADRLVYHGDPAVWVWSQVVKTLIRPQPLEKEIEBARKL 353
DB 306 AVEELHKLKKEAGEANKAIOHLLLEVEQSK-----DQMEKEMLEKIGRLKELELEMANDLL 360
OY 354 GFKHPIVIGVHVRITDKVGA-----EAAFPH-----IREYTVHVEEDFOL 392
DB 361 S-----ATKRKCAIILSEELAMSPPTAAAVAKIVKPKMLTELYNAVYETODOL 409
OY 393 LARRMOYDKKRVYLATDPPALLKEAKTKYP 422
DB 410 LLEKL--ENKRINKYLDL--IVKEVEAKAP 435

RESULT 3
MYSD_CAEEL STANDARD: PRT: 1938 AA.
AC P02567: 019674:
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain D (MHC D).
GN MYO-1 OR R06C7.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE-89178677; PubMed-2926820;
RA Dobb N.J., Maruyama I.N., Krause M., Karn J.;
RT "Sequence analysis of the complete Caenorhabditis elegans myosin
heavy chain gene family."
RL J. Mol. Biol. 205:603-613(1989).
RN 12
RP SEQUENCE OF 34-1795 FROM N.A.
RX MEDLINE-83273600; PubMed-6576334;
RA Karn J., Brenner S., Barnett L.;
RT "Protein structural domains in the Caenorhabditis elegans unc-54
myosin heavy chain gene are not separated by introns."
RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
RN 13
RP SEQUENCE OF 115-365 AND 1492-1763 FROM N.A.
RX MEDLINE-85201409; PubMed-3888374;
RA Karn J., Dobb N.J., Miller D.M.;
RT "Cloning nematode myosin genes."
RL Cell Muscle Motil. 6:185-237(1985).
RN 14
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gardner A., McMurray A.;
RL submitted (APR-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
AND 2 REGULATOR LIGHT CHAIN SUBUNITS (RLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.

```

```

CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES.
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PPM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
C.ELEGANS.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb.ch).
CC -----
DR EMBL: X08065; CAA30854.1; -
DR EMBL: M37232; AAA28119.1; -
DR EMBL: M37234; AAA28120.1; -
DR EMBL: Z71266; CAA95848.1; -
DR EMBL: Z71261; CAA95805.1; JOINED.
DR EMBL: Z71266; CAA95805.1; JOINED.
DR PIR: S02772; MMKWI.
DR HSSP: P08799; IMND.
DR WormPep: R06C7.10; CE06253.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_Tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head.1.
DR Pfam: PF02736; Myosin_N.1.
DR Pfam: PF01576; Myosin_Tail.1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; MYOSINHEAD.1.
DR SMART: SM00242; MYSC.1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
ATP-binding; Methylation; Alkylation; Multigene family.
FT DOMAIN 845 1938
FT DOMAIN 846 1170
FT DOMAIN 1171 1938
FT DOMAIN 846 1938
FT NP_BIND 177 184
FT DOMAIN 660 682
FT DOMAIN 764 778
FT MOD_RES 128 128
FT MOD_RES 700 700
FT MOD_RES 710 710
FT MOD_RES 94 94
FT CONFLICT 94 98
FT CONFLICT 98 98
FT CONFLICT 377 377
FT CONFLICT 389 390
FT CONFLICT 391 391
FT CONFLICT 408 408
FT CONFLICT 474 474
FT CONFLICT 577 577
FT CONFLICT 681 681
FT CONFLICT 1373 1373
FT CONFLICT 1659 1659
SQ SEQUENCE 1938 AA; 223255 MW; 387399C8F63AACF4 CRC64;
Query Match 3.8%; Score 117; DB 1; Length 1938;
Best Local Similarity 20.5%; Pred. No. 3.7;
Matches 97; Conservative 70; Mismatches 139; Indels 168; Gaps 21;
OY 36 SDHSSRELSTKLAKLERKQONEDLR--MAGSLRIPECPIDOGPASGR-----V 83
DB 1481 SRNTSTVEFKLRSMWINDSEQIETLRRENKIFSGEIRINDIQOG--GRTYQEVHKS 1537

```


[4]
RP SEQUENCE OF 739-1074 FROM N.A. (ISOFORM PLD1D).
RA Hughes W.E., Parker P.J.;
RT "A novel human phospholipase D1 splice variant displays conserved regulation in vitro but altered localisation in vivo."; submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP TISSUE SPECIFICITY.
RX MEDLINE=98250727; PubMed=958213;
RA Lopez I., Arnold R.S., Lambeth J.D.;
RT "Cloning and initial characterization of a human phospholipase D2 (hPLD2). ADP-ribosylation factor regulates hPLD2."; J. Biol. Chem. 273:12846-12852(1998).
RL
CC -1- FUNCTION: IMPLICATED AS A CRITICAL STEP IN NUMEROUS CELLULAR PATHWAYS, INCLUDING SIGNAL TRANSDUCTION, MEMBRANE TRAFFICKING, AND THE REGULATION OF MITOSIS. MAY BE INVOLVED IN THE REGULATION OF PERINCULAR INTRAVASCULAR MEMBRANE TRAFFIC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: A phosphatidylcholine + H(2)O -> choline + a phosphatidate.
CC -1- ENZYME REGULATION: STIMULATED BY PHOSPHATIDYLINOSITOL 4,5-BISPHOSPHATE AND PHOSPHATIDYLINOSITOL 3,4,5-TRISPHOSPHATE, ACTIVATED BY THE PHOSPHOKINASE C-ALPHA, BY THE ADP-RIBOSYLATION FACTOR-1 (ARF-1), AND IN A LESSER EXTENT BY GTP-BINDING PROTEINS: RHO A, RAC-1 AND CDC42. INHIBITED BY OLEATE.
CC -1- SUBCELLULAR LOCATION: PERINUCLEAR REGIONS; ENDOPLASMIC RETICULUM, GOLGI APPARATUS AND LATE ENDOSOMES (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: PLD1A (SHOWN HERE), PLD1B, PLD1C AND PLD1D; ARE PRODUCED BY ALTERNATIVE SPLICING. PLD1C IS UNLIKELY TO BE PHYSIOLOGICALLY RELEVANT.
CC -1- TISSUE SPECIFICITY: EXPRESSED ABUNDANTLY IN THE PANCREAS AND HEART AND AT HIGH LEVELS IN BRAIN, PLACENTA, SPLEEN, UTERUS, AND SMALL INTESTINE.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE D FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PHOX HOMOLGY (PX) DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 PLDC DOMAINS.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce>).
CC or send an email to license@isb-sib.ch.
CC
CC EMBL: U38545; AAB49031.1; -.
CC EMBL: AJ276230; CAB76564.1; -.
CC
DR MIM: 602382; -.
DR InterPro: IPR001849; PH.
DR InterPro: IPR001736; PL.
DR InterPro: IPR001683; PX.
DR Pfam: PF00169; PH: 1.
DR Pfam: PF00614; PLDC: 2.
DR Pfam: PF00787; PX: 1.
DR SMART: SM00233; PH: 1.
DR SMART: SM00155; PLDC: 2.
DR SMART: SM00312; PX: 1.
DR PROSITE: PSS00003; PH DOMAIN: FALSE_NEG.
KW Hydrolase; Lipid degradation; Membrane; Alternative splicing;
KW Endoplasmic reticulum; Golgi stack; Repeat.
FT DOMAIN 76 196
FT DOMAIN 219 328
FT DOMAIN 459 486
FT DOMAIN 891 918
FT DOMAIN 463 928
FT VARSPLIC 514 597
FT
FT PAAHSEMSDLSLKKRNPEVQNLPIQKSIDVDYDKLKGIGKP
FT RKFSEFLYKQDLHHHLDADISISIDSTSSYFNMHSHNH
FT LI -> IPGEVVYRVQWSECMGRPDSGMETFTAISSSKTG
FT FNLNLLISLITGPRPCGCMPTLPLOSTGRRLVMMHVTSS
FT SAGTSOKI (IN ISOFORM PLD1C).
FT MISSING (IN ISOFORM PLD1D).
FT SYFNHYSHNLHGLKRPHEKLFHPSSSESGGLTRPHAD
FT -> N (IN ISOFORM PLD1B).
FT
FT VARSPLIC 598 1074
FT VARSPLIC 585 623

FT	VARSPLIC	962	971	VTGLYDDPS -> SKMTPEVEDP (IN ISOFORM PLDID).
FT	VARSPLIC	972	1074	MISSING (IN ISOFORM PLDID).
FT	CONFLICT	832	832	S -> P (IN REF. 3).
SO	SEQUENCE	1074 AA:	124184 MW:	734F285790ADBE7A CRC64;
Query Match 3.64; Score 110.5; DB 1; Length 1074;				
Best Local Similarity 18.04; Pred. No. 4.9;				
Matches 101; Conservative 67; Mismatches 166; Indels 227; Gaps				
OY	114 ILRRRIENGAK-----ELMFLESELEK--LKNLEGG-ELORADEFLSDL--GH 158			
DB	404 ILKRAAQGVAFRLMYLKYKEVELAGINSEYKTKRLMLRHPKIKMKRPRDHVSSTVILMAH 463			
OY	159 HERSIMTD--LYULSQTDGA--GDWREKAKDLTELVOIRTYLQNPKDCSAKKLYCININ 215			
DB	464 HEKLIYIDQSAFAFGIDLAIVGRMDNE-----HRLIDVGSVAKRVISGPEL----- 509			
OY	216 KGGCGCOLHHVYCEMIAVGTORTLALESNNMRYATGGMETVFRPVSECTDRSGSSTG 275			
DB	510 -----GSLPRAMESME----- 521			
OY	276 HMSGVKKKNQYVELPT-----VDS----- 296			
DB	522 --SLRLAKDKNEPVOMPIQKSIDVDYSLKIGKPRKFSKSLYKOLHRHLLHADSISS 579			
OY	297 -----VHPRPYRLPLAVPELDLDR--LVR-----VHG 321			
DB	580 IDSTSYSTNNHRSNNHLLHGKLRPKLFHPSESEOGITRNADGTSIRLSQTGVGELHG 639			
OY	322 DPAVWVWSOFVKYLLIRPQWLEKEIEE----ATKKLGFKHPIGVHVRRTDKVGAABAF 376			
DB	640 ETRFNHGKDYCNFVFKDWOLDKPFADFDIRSTRPMRMDIASAVH---GKARPVAR 695			
OY	377 HPLEETYLVHVEDRQLARRQVQDKKRYLLATDDPALLKEAKTKYRPERFISDMSI---- 432			
DB	696 HFTQW-----NFKTI-----MKSXYRSLSYPLPKSQTANHELRYOVPSVIANV 742			
OY	433 -----SMSAGLNHRYENSLRGVILDI-----HFLSADFLVCFSSQVC 472			
DB	743 QLLRRAADMSKGI--KYHEESIHAAVYVENSRIHYIENQFISCADKV--VFNKIGD 799			
OY	473 RYAVEYIMQA-----LHPDASANFRSLDIDYFGGNANHOIAIYRPHRTE 518			
DB	800 AIAQRIILAHRENQRYRVYVYIPLRPGEGDIST-----GGNALQAIWHFNRYRTMCR 852			
OY	519 GEIPMEPGDIIIGVAGNMHDGY 539			
DB	853 GENSI-LGOLKAELGNOMINY 872			
RESULT 9				
SCPL1_HUMAN				
ID	SCPL1_HUMAN	STANDARD;	PRT;	976 AA.
AC	Q15431; Q14963;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Synaptonemal complex protein 1 (SCP-1 protein).			
GN	SCPL1 OR SCPL1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Testis;			
RX	MEDLINE=97224467; PubMed=9119375;			
RA	Meuwissen R. L.J., Meerts I., Hoovers J. M. N., Leeschot N. J.,			
RT	"Human synaptonemal complex protein 1 (SCP1): Isolation and			
RT	characterization of the cDNA and chromosomal localization of the			
RT	gene";			

RL Genomics 39:377-384(1997).
 RN [2]
 RC SEQUENCE FROM N.A.
 RX MEDLINE=98037449; PubMed=9371398;
 RA Kondoh N., Nishina Y., Tsuchida J., Koga M., Tanaka H., Uchida K.,
 RA Inazawa J., Taketo M., Nozaki M., Nojima H., Matsunaga K., Namiki M.,
 RA Okuyama A., Nishimune Y.;
 RT "Assignment of synaptonemal complex protein 1 (SCP1) to human
 RT chromosome 1p33 by fluorescence in situ hybridization and its
 RT expression in the testis";
 RL Cytogenet. Cell Genet. 78:103-104(1997).
 CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: TESTIS.
 CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
 CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
 CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL: X95654; CAA64956.1; -
 DR EMBL: D67035; BAA22586.1; -
 DR MIM: 602162; -
 KW Nuclear protein; Meiosis; Cell division; Phosphorylation;
 KM DNA-binding; Coiled coil.
 FT DOMAIN 107 120
 FT 117 120 ASP/GLU-RICH (ACIDIC).
 FT 679 682 COILED COIL (POTENTIAL).
 FT 880 883 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 961 969 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT 46 46 ARG/LYS-RICH (BASIS).
 FT 106 106 L -> F (IN REF. 2).
 FT 106 106 F -> Y (IN REF. 2).
 FT 153 153 F -> C (IN REF. 2).
 FT 161 161 K -> T (IN REF. 2).
 FT 168 168 E -> D (IN REF. 2).
 FT 216 216 N -> S (IN REF. 2).
 FT 225 226 HG -> FE (IN REF. 2).
 FT 350 350 K -> N (IN REF. 2).
 FT 360 360 E -> D (IN REF. 2).
 FT 400 400 KN -> NY (IN REF. 2).
 FT 406 406 K -> I (IN REF. 2).
 FT 415 415 K -> T (IN REF. 2).
 FT 449 449 E -> D (IN REF. 2).
 FT 483 510 TOLTAITSEQYYSKVEKLTKELENEK -> YSYCHYHKM
 FT 510
 FT 516 528 TVLPKRGPKLSKRE (IN REF. 2).
 FT 528
 FT 549 549 LTSHCKKLSLEKN -> YFTLOASPPN (IN REF.
 FT 560 560 2).
 FT 805 805 N -> I (IN REF. 2).
 FT 941 941 K -> T (IN REF. 2).
 FT 941 941 E -> D (IN REF. 2).
 FT 941 941 P -> S (IN REF. 2).
 SQ SEQUENCE 976 AA; 114069 MW; 8BA8ID042AC2696B CRC64;
 Query Match 3.6%; Score 110; DB 1; Length 976;
 Best local similarity 20.5%; Pred. No. 4.7;
 Matches 83; Conservative 69; Mismatches 152; Indels 100; Gaps 16;

Db 496 SKEVKDLTELENEKLNKLTSHCKKLSLENNKELTQ-ETSDMTLELKNQ----QEDINN 550
 Qy 100 YKKQTKNGCKKHHELLRRIENGAKELMFFLOSELKUKLNKNGNELQRIADFEGLDGH 159
 Db 551 NKKQ-----EENLKQIEN-LOETQTRNELEYR---EELKKORDEVKCKL--- 594
 Qy 160 ERSIMTDLXYLSQTDGAGMKREKAKDLTELVRRTTYLQNPDKCSKAKKLTGNKCGC 219
 Db 595 -----DKSEENCNLLKQYENKNNKYIEELQENKALK-----KKQTA 631
 Qy 220 YCCQLHHVYCFMAYGTOFTLALESHNRVATGNETVFRPVSETCIDRSSSTGHSNG 279
 Db 632 ESKQLN--VEYELKV--NKLELELESARKQF-----GEITDTYOK 666
 Qy 280 EYKDKVQVVELPIYDSVHPRPYLPLAVPELDADLVAVHGD---PAVWWSQFYKYL 336
 Db 667 ELEDKRI--EENLLEVEK-----AKVIADAVVLQKEIDKRCQKTAENVALME 715
 Qy 337 RPOWLEKEIEEATKRLGKHNPIGVHVRPTDVGAEAFHPIEETVAVEDFOLLAR 396
 Db 716 KKHQYDKLIEERDSEL-----GLYKSEDOSSLRSLELSNLKAE---LLSVK 764
 Qy 397 KQVD-----KKRYLATDPLLKEAKTKYPSYEFISONSISW 434
 Db 765 KOLEIEREKEKRLBREKENTATLKEKKDKQTQTELETPEIYW 808
 RESULT 10
 SCPL_MOUSE
 ID SCPL_MOUSE STANDARD; PRT: 993 AA.
 AC Q62209; Q62329; Q09205; P70192;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Synaptonemal complex protein 1 (SCP-1 protein).
 GN SCPL OR SCPL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBA; TISSUE=Testis;
 RX MEDLINE=96004899; PubMed=7548215;
 RA Sage J., Martin L., Cuzin F., Rassoulzadegan M.;
 RT "cDNA sequence of the murine synaptonemal complex protein 1 (SCP1).";
 RL Biochim. Biophys. Acta 1263:258-260(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SWISS; TISSUE=Testis;
 RA Kerr S.M., Taggart M.H., Lee M., Cooke H.J.;
 RL Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-149 FROM N.A.
 RC STRAIN=C57BL/6;
 RA Sage J., Li Y., Martin L., Matel M.-G., Guenet J.-L., Liu J.G.,
 RA Hoeg C., Cuzin F., Rassoulzadegan M.;
 RL Submitted (Jan-1997) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 95-787 FROM N.A.
 RC STRAIN=ICR; TISSUE=Testis;
 RA Tsuchida J., Nishina Y., Nozaki M., Uchida K., Nishimune Y.;
 RL Submitted (Nov-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MAJOR COMPONENT OF THE TRANSVERSE FILAMENTS OF
 CC SYNAPTONEMAL COMPLEXES (SCS), FORMED BETWEEN HOMOLOGOUS
 CC CHROMOSOMES DURING MEIOTIC PROPHASE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. IN TRIPARTITE SEGMENTS OF
 CC SYNAPTONEMAL COMPLEXES, BETWEEN LATERAL ELEMENTS IN THE NUCLEUS.
 CC FOUND ONLY WHERE THE CHROMOSOME CORES ARE SYNAPSED. ITS N-TERMINUS
 CC IS FOUND TOWARDS THE CENTRE OF THE SYNAPTONEMAL COMPLEX WHILE THE
 CC C-TERMINUS EXTENDS WELL INTO THE LATERAL DOMAIN OF THE
 CC SYNAPTONEMAL COMPLEX (BY SIMILARITY).
 CC -----

CC -1- DOMAIN: CONSISTS OF AN ALPHA-HELICAL STRETCH OF 700 AA RESIDUES,
CC FLANKED BY N- AND C-TERMINAL GLOBULAR DOMAINS. THE C-TERMINAL
CC DOMAIN HAS DNA-BINDING CAPACITY (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 238118; CAA86262.1; -
DR EMBL: U41069; AAA64514.1; ALT_INT.
DR EMBL: U62864; AAC53335.1; -
DR EMBL: U62860; AAC53335.1; JOINED.
DR EMBL: U62861; AAC53335.1; JOINED.
DR EMBL: U62862; AAC53335.1; JOINED.
DR EMBL: U62863; AAC53335.1; JOINED.
DR EMBL: D88539; BAA13639.1; -
DR MGD: MGI:105931; Sycepl.
KW Nuclear protein; Melosins; Cell division; Phosphorylation;
KW DNA-binding; Coiled coil.
FT DOMAIN 12 97 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 104 815 COILED COIL (POTENTIAL).
FT DOMAIN 114 117 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 697 700 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 898 901 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 978 986 ARG/LYS-RICH (BASIC).
FT CONFLICT 527 527 F -> L (IN REF. 2).
SQ SEQUENCE 993 AA; 115962 MW; 1A4FA790D64FAFE6 CRC64;

Query Match 3.6%; Score 110; DB 1; Length 993;
Best Local Similarity 18.9%; Pred. No. 4.8;
Matches 85; Conservative 78; Mismatches 175; Indels 112; Gaps 17;

OY 42 ELSTLAKLERKQONEDRRMAGSLRIPEPID--QGPASRVRLAEQFKAKEQIEN 99
DB 432 ELKNTLAEQKLLDEKQYKLAELQEKQELTFLPREKREVIDLQVYVYTSSEH 491
OY 100 YKQTRNGPGKDEILRRRIENGAKELWFLQSELKILN-----LEGNELORH 148
DB 492 YLKQ-----VEEMKTEL-----EKELKNTELTASCDMLLENKKFVGE 530
OY 149 ADEFLSDGCHHRSTMT---DLYIISOTDAGDMREKAKDLTELVORRYTLQNPKD- 203
DB 531 ASDMALTELKHOEDIIKCKKOEKRLKOLE---NLEEKEMHLRDELSVRKEFIQOGDEV 587
OY 204 -----CSKAKKILVCNINKGGYGCOLHHVYCFM 232
DB 588 KKKLDSSEMASTICEVUKKEKOKILKSCNNLKKOVENSKNIE---ELHQ----- 638
OY 233 IAYGQRTLALSHMWRVATGMEVFRPV--ETCTDRSGSSGTGHMSGEVKNVQV 289
DB 639 -----ENKTLKKSSAEIKOLNAYEIKVSKLELESTKQRFPEEMTNVNYKEIENKIS-- 692
OY 290 ELPIYDSVHPRPYPLAVPEDLADRLVAVHGDPAVMWVSQVYKILIRPQPLEKEIEEA 349
DB 693 ECKLLGEVEKAKATYDEAKVLOKEIDLRCQHK-----IAENVALMEKHKHOYDIIVER 746
OY 350 TKLLG-FKHPIVGVHVRRTDKVGAEAAPHIEEYVHVEDEDQLARRQV---DKKRYV 405
DB 747 DSELGLYKN-----REEGSSAKTA---LETLSNINNELVSLKKQLEIKKEKEKELK 796
OY 406 LATDDPALKEAKTYRPSYEFT-SDNSISW 434
DB 797 MAKENTALIKDKKKIKQASLSLESPATSW 826

RESULT 11
CNF_HUMAN STANDARD: PRT; 3210 AA.
ID CNF_HUMAN
AC P49454; Q13246; Q13171;

DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosin) (AH
DE antigen).
GN CENP-F.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Breast carcinoma;
RX MEDLINE=953348175; PubMed=7542657;
RA Liao H., Winkfeld R.J., Mack G., Rattner J.B., Yen T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto
RT kinetochores at late G2 and is rapidly degraded after mitosis";
RL J. Cell Biol. 130:507-518(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B.,
RA Jones D., Yang-Feng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein
RT that is specifically involved in mitotic-phase progression";
RL Mol. Cell. Biol. 15:5017-5029(1995).
RN [3]
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=95336446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Partig N., Medsker T.A., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal
RT domain sufficient for nuclear localization";
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization,
RT centromere/kinetochore targeting, and dimerization";
RL J. Biol. Chem. 270:19545-19550(1995).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1";
RL J. Cell Biol. 143:49-63(1998).
CC -1- FUNCTION: PROBABLY REQUIRED FOR KINETOCORE FUNCTION, INVOLVED IN
CC CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH
CC RETINOBLASTOMA PROTEIN (RB), CENP-E AND BUBR1.
CC -1- SUBUNIT: HOMO- OR HETERO-DIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS),
CC REORGANIZATION TO THE KINETOCORE/CENTROMERE (CORONAL SURFACE OF
CC THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
CC -1- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
CC -1- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U19769; AAA82889.1; -
DR EMBL: U30872; AAA82935.1; -
DR EMBL: U25725; AAA86839.1; -
DR HSP: P02649; ILE4.
DR MIM: 600236; -
KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
FT DOMAIN 14 197 COILED COIL (POTENTIAL).

```
FT DOMAIN 273 769 COILED COIL (POTENTIAL).
FT DOMAIN 823 1328 COILED COIL (POTENTIAL).
FT DOMAIN 1642 1746 COILED COIL (POTENTIAL).
FT DOMAIN 1862 2987 COILED COIL (POTENTIAL).
FT DOMAIN 2207 2560 2 X 177 AA TANDEM REPEATS.
FT REPEAT 2207 2386 1.
FT REPEAT 2389 2568 2.
FT DOMAIN 3032 3532 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 16 16 T -> A (IN REF. 2).
FT CONFLICT 250 250 L -> Q (IN REF. 2).
FT CONFLICT 272 272 G -> D (IN REF. 2).
FT CONFLICT 611 611 MISSING (IN REF. 2).
FT CONFLICT 1494 1589 MISSING (IN REF. 2).
FT CONFLICT 1611 1611 V -> A (IN REF. 2).
FT CONFLICT 1811 1811 V -> L (IN REF. 2).
FT CONFLICT 2242 2243 ER -> DG (IN REF. 3).
FT CONFLICT 2335 2335 L -> Q (IN REF. 3).
FT CONFLICT 2492 2492 D -> N (IN REF. 2).
FT CONFLICT 2545 2561 ELNERVALHNDQECK -> SSREMPQPCIMTKRVS
      (IN REF. 3).
SQ SEQUENCE 3210 AA; 367589 MM; 11D83324960E4334 CRC64;

Query Match 3.6%; Score 110; DB 1; Length 3210;
Best Local Similarity 19.1%; Pred. No. 23;
Matches 99; Conservative 86; Mismatches 190; Indels 142; Gaps 24;
```

```
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxId=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kertavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weisslock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC diphosphate + L-phenylalanyl-tRNA(Phe).
CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (by
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-tRNA SYNTHETASE FAMILY.
CC PHE-tRNA SYNTHETASE ALPHA CHAIN SUBFAMILY 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC
CC EMBL: U67499; AAB98478.1; -
CC HSP: P27001; 1PVS.
CC TIGR: M0487; -
CC InterPro: IPR002106; AA_tRNA_ligase_II.
CC InterPro: IPR002319; tRNA-synt_2d.
CC Pfam: PF01409; tRNA-synt_2d.1.
CC PROSITE: PS00179; AA_tRNA_ligase_II_1.1.
CC DR PROSITE: PS00339; AA_tRNA_ligase_II_2; FALSE_NEG.
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
CC KW Complete proteome.
SQ SEQUENCE 480 AA; 57027 MM; 1DC2599CF0D97B72 CRC64;

Query Match 3.5%; Score 109.5; DB 1; Length 480;
Best Local Similarity 18.4%; Pred. No. 2;
Matches 105; Conservative 84; Mismatches 159; Indels 223; Gaps 28;
```

DB 230 -----LTRIR-----EVEKILLANGFEKVKSP 252
OY 359 VIGVHVRTRDKVGAFAHFEIETVHVEDPQLLARBMQVOKRRVYLATDPALEK 418
DB 253 IVETEFMNFDM-L-PEPODH-----AREMO-----DTPFL-----281
OY 419 TTPSYEFISDINSISMSACLHNR-----YTENSLRGVILIDHFLSQADEFLVCTFSOYCR 473
DB 282 -KYPNEGDIPELLLSKVYKHERCKMKYKDENVSRLLRTTTTASSINYLASLSDEEKN 340
OY 474 VAYET-MOALHPDASANFESLDDIYFEGC-----PNAHNOI-----508
DB 341 KEHVFCEIDRVFRNEAIDKHLPEFYOCGCIIMDDNVFNENLIGVLKEFLNLFGEKVAF 400
OY 509 --AIVPH-OPTREGELPMPEGDIIGVAGNHM 536
DB 401 RPAVFEFTPEPSLEAEVYLE-----GKGW 423

RESULT 13
OSFL_HUMAN
ID OSFL_HUMAN STANDARD: PRT: 214 AA.
AC 092882;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Osteoclast stimulating factor 1.
CN OSTF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RA MEDLINE=99190637; PubMed=10092216;
RA Reddy S.V., Devlin R., Mena C., Nishimura R., Choi S.J., Dallas M.,
RA Yoneda T., Roodman G.D.;
RT "Isolation and characterization of a cDNA clone encoding a novel
peptide (OSF) that enhances osteoclast formation and bone
resorption.";
RT J. Cell. Physiol. 177:636-645(1998).
RN [2]
RP SMN1 BINDING.
RA Roodman G.D.;
RL Unpublished observations (NOV-2000).
CC -1- FUNCTION: INDUCES BONE RESORPTION, ACTING PROBABLY THROUGH A
CC SIGNALING CASCADE WHICH RESULTS IN THE SECRETION OF FACTOR(S)
CC ENHANCING OSTEOCLAST FORMATION AND ACTIVITY.
CC -1- SUBUNIT: BINDS C-SRC, AND SMN1.
CC -1- TISSUE SPECIFICITY: URINARYLY EXPRESSED.
CC -1- SIMILARITY: CONTAINS 3 SH3 REPEATS.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U63717; AAB06396.1; -
DR HSSP: P06241; ISHF.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001452; SH3.
DR Pfam: PF000023; ank; 3.
DR Pfam: PF00018; SH3; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART: SM00248; ANK; 2.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS50002; SH3; 1.

DR PROSITE: PS50297; ANK_REPEAT; 1.
DR PROSITE: PS50088; ANK_REPEAT; 1.
KW SH3 domain; Repeat; ANK repeat.
FT DOMAIN 4 11 PRO-RICH.
FT DOMAIN 12 71 SH3.
FT REPEAT 72 101 ANK 1.
FT REPEAT 105 135 ANK 2.
FT REPEAT 139 168 ANK 3.
SQ SEQUENCE 214 AA: 23799 MW: E10C6570B68E3D6D CRC64:
Query Match 3.5%; Score 109; DB 1; Length 214;
Best Local Similarity 38.0%; Pred. No. 0.74;
Matches 27; Conservative 9; Mismatches 25; Indels 10; Gaps 2;
OY 509 AIVPHOPTREGELPMPEGDIIGVAG-----NHWDGTPKGVNKRGLTGLPSKYVKEKET 564
DB 19 ALVTEPRTPELDYFEEGDIIVTDMSDTNMKGTSK-----GRTGLPSVYVAQAES 72
OY 565 VKPTYPEADK 575
DB 73 IDNPLHEAKR 83

RESULT 14
NODZ_BRAJA
ID NODZ_BRAJA STANDARD: PRT: 324 AA.
AC 045271;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Nodulation protein 2 (EC 2.4.1.-).
GN NODZ.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RA MEDLINE=94131940; PubMed=8300517;
RA Stacey G., Luka S., Sanjuan J., Bantlavi Z., Nieuwkoop A.J.,
RA Chun J.Y., Forsberg L.S., Carlson R.;
RT "nodZ, a unique host-specific nodulation gene, is involved in the
RT fucosylation of the lipooligosaccharide nodulation signal of
RT Bradyrhizobium japonicum.";
RL J. Bacteriol. 176:620-633(1994).
CC -1- FUNCTION: FUCOSYLTRANSFERASE WHICH ADDS THE FUCOSE MOIETY OF THE
CC NOD FACTOR ON ITS TERMINAL REDUCING N-ACETYLGLUCOSAMINE END. USES
CC GDP-FUCOSE AS THE DONOR GROUP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L22756; AAB58613.1; -
KW Nodulation; Transferase; Glycosyltransferase.
SQ SEQUENCE 324 AA: 36668 MW: 0DA83CFB9E040E22 CRC64:
Query Match 3.5%; Score 109; DB 1; Length 324;
Best Local Similarity 21.2%; Pred. No. 1.3; Mismatches 10; Indels 14;
Matches 68; Conservative 48; Mismatches 10; Indels 14;
OY 182 KEAKDLTELVRRTIYIQPKDSKAKKLVNCIN-----KGGYGCQLHHVYVCFWI 233
DB 23 RETSVLTSLVQ-----FGAREKARQGVSSGSSNDPRFVSSRRRGFGGCLSLAAARNF 74
OY 234 AVGTORTALESNMRYA-----TGQMETVFRPVSE-----TCTDRSGSSTGHWGCE 280
DB 75 AKOTGRTLAID--WRGSCYLDPEPTNAPVPEFVEDIGGVVICDD-----119

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 31, 2002, 13:07:01 : Search time 22 seconds
(without alignments)
2511.426 Million cell updates/sec

Title: US-09-839-136-2
Perfect score: 3096
Sequence: 1 MRPWGSMRWIMLILFAWGT.....YKVKREIETVKYPPYPPADK 575

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2949	95.3	575	1 JCS432	glycoprotein 6- α p
2	948	30.6	818	2 T32154	hypothetical prote
3	140	4.5	1025	2 H86250	hypothetical prote
4	123.5	4.0	1057	2 F89892	cardamoyl-phosphat
5	123	4.0	1871	2 D96796	probable heat shoc
6	122.5	4.0	1427	2 S22695	reslin - human
7	121.5	3.9	1154	2 F71856	hypothetical prote
8	118.5	3.8	1392	2 M43336	microtubule-vesicl
9	117.5	3.8	2094	2 S33124	lpr protein - huma
10	117	3.8	1938	1 MKWK1	myosin heavy chain
11	116	3.7	544	2 T32568	hypothetical prote
12	114.5	3.7	845	2 I48176	synaptonemal compl
13	112	3.6	978	2 H89971	conserved hypothet
14	111.5	3.6	978	2 A70387	conserved hypothet
15	111.5	3.6	1071	2 F39845	cardamoyl-phosphat
16	111	3.6	398	2 T04501	hypothetical prote
17	111	3.6	570	2 C90485	beta-glucuronidase
18	111	3.6	1011	2 T13055	dynamn associated
19	111	3.6	1094	2 T13053	dynamn associated
20	111	3.6	1549	2 T21809	hypothetical prote
21	111	3.6	1940	2 A59287	myosin heavy chain
22	110.5	3.6	909	2 T00009	probable primase (
23	110	3.6	993	2 S49461	synaptonemal compl
24	109.5	3.5	480	2 G64360	phenylalanine--tRN
25	109.5	3.5	1179	2 F71190	probable chromosom
26	109	3.5	324	2 A36959	host-specific nodu
27	109	3.5	417	2 D75183	hypothetical prote
28	109	3.5	822	2 G69010	conserved hypothet
29	108.5	3.5	1510	2 T16927	hypothetical prote

30	108.5	3.5	1738	2 T14867	interaplin - slime
31	108	3.5	678	2 H88187	protein C18H9.8 (1
32	107.5	3.5	503	4 S23741	hypothetical trp/T
33	107.5	3.5	2288	2 T29999	hypothetical prote
34	107	3.5	2441	2 D71623	erythrocyte membra
35	106.5	3.4	900	2 T19689	hypothetical prote
36	106.5	3.4	2326	2 T29140	hypothetical prote
37	106	3.4	581	2 B87768	protein P54C1.1 (1
38	106	3.4	1992	1 S02771	myosin heavy chain
39	105.5	3.4	487	1 T32635	phosphoprotein pho
40	105.5	3.4	646	2 A44985	dnak-type molecula
41	105.5	3.4	758	2 H87133	methylnalonyl-CoA
42	105.5	3.4	1188	2 E89896	chromosome segrega
43	105	3.4	674	1 DAHUAL	archidonate 5- β -lip
44	105	3.4	880	2 F75102	conserved hypothet
45	105	3.4	919	2 B72765	hypothetical prote

ALIGNMENTS

RESULT 1

JCS432

glycoprotein 6- α -L-fucosyltransferase (EC 2.4.1.68) precursor - human

N:Alternate names: alpha 1-6 FucT; GDP-L-fucose:N-acetyl-beta-D-glucosaminyl:6- α -

C:Species: Homo sapiens (man)

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Jun-2000

C:Accession: JCS432; PC4322

R:Yanagidani, S.; Uozumi, N.; Ihara, Y.; Miyoshi, E.; Yamaguchi, N.

J. Biochem. 121, 626-632, 1997

A:Title: Purification and cDNA cloning of GDP-L-Fuc:N-acetyl-beta-D-glucosaminide:alp

A:Reference number: JCS432; MUID:97279058

A:Accession: JCS432

A:Molecule type: mRNA

A:Residues: 1-575 <YAN1>

A:Cross-References: DDBJ: D89289; NID: g2055306; PID: BAA19764.1; PID: g2055307

A:Accession: PC4322

A:Molecule type: protein

A:Residues: 68-87:352-376:419-432 <YAN2>

C:Comment: This enzyme catalyzes the transfer of fucose from GDP-fucopyranoside to as

C:Genetics:

A:Gene: GDB: FUT8

A:Cross-References: GDB: 9786294; OMIM: 602589

A:Map position: 14q23.1q23

C:Superfamily: human glycoprotein 6- α -L-fucosyltransferase

C:Keywords: glycosyltransferase; hexosyltransferase

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-575/Product: glycoprotein 6- α -L-fucosyltransferase #status predicted <MAT>

Query Match 95.3% Score 2949; DB 1; Length 575;

Best Local Similarity 95.5%; Pred. No. 1.3e+197; Mismatches 549; Conservative 6; Mismatches 20; Indels 0; Gaps 0;

QY	1	MRPWGSMRWIMLILFAWGTLLFYIGHLVLRNDSDHSRSREISKLAKLEKQONEDL 60		
DB	1	MRPWGSMRWIMLILFAWGTLLFYIGHLVLRNDSDHSRSREISKLAKLEKQONEDL 60		
QY	61	RMASGLRIPEDIPDGPASGRVRALEDFPMKAKQEIENYKKQTKNGPKDHEILRRRIE 120		
DB	61	RMASGLRIPEDIPDGPASGRVRALEDFPMKAKQEIENYKKQTKNGPKDHEILRRRIE 120		
QY	121	NAKELMPFLOSELKKKLKLEGNELORHADFSLDIGHHERSIMDLYLSTODGADMR 180		
DB	121	NAKELMPFLOSELKKKLKLEGNELORHADFSLDIGHHERSIMDLYLSTODGADMR 180		
QY	181	EKEADLTELVRITTYLQNPDKCSKAKKLVNINKGCGYGCQLHHVVYCFMIAVGTQRT 240		
DB	181	EKEADLTELVRITTYLQNPDKCSKAKKLVNINKGCGYGCQLHHVVYCFMIAVGTQRT 240		
QY	241	LALSHNNRYATGKEVYFRPVSEFTCDRSGSTGHWSEVYDKNVQVELPIVDSVAPR 300		
DB	241	LALSHNNRYATGKEVYFRPVSEFTCDRSGSTGHWSEVYDKNVQVELPIVDSVAPR 300		

Db 736 NLEFGANAGKAMTFADKATVSPYAKEVAGNSYSAHLKFKHGIINGIDPIDIDPND 795
QY 330 QFVYTLRPOWLEKEI-----EEATKLGFK---HPYGVNVRTDKVGAARAHNP 378
Db 796 NFI-----PVPTYSBNVEGKRAKEELONRGKLSADFPVVGIIITRLTHOKGHLIKHA 850
QY 379 I-----EEYTVHEEDFOLLARRMOV---DKRVYLATDPPALLKFAKTRY 421
Db 851 IWRTERNGOYVLLGASADPRQNDNFVNLQNLHSHSGDRARLVLTDEPL-----SHLTY 906
QY 422 PSYEFISDNISWSAGL 438
Db 907 AGADFIIVPSIFEPCGL 923

RESULT 4
F89692
cardamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: F89692
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: F89692
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1057 <KUR>
A:Cross-references: GB:BA000018; PID:q13701002; PIDN:BAB42298.1; GSPDB:GN00149
A:Experimental source: strain N315
A:Gene: PYAB
C:Superfamily: cardamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin

Query Match 4.0%; Score 123.5; DB 2; Length 1057;
Best Local Similarity 18.9%; Pred. No. 2;
Matches 115; Conservative 85; Mismatches 221; Indels 189; Gaps 25;
QY 43 LSKILAKLERLKOEDLRMMAGS-----LRPEGPID-----OGP 78
Db 318 IAKLAIAVGLTDEMNPITGSTYAFEPFLDVIKIRPFPRDKKEKRELGSTQM 377
QY 79 ASGRVAL-----EEOFMKAKEQIEN---YKQTKNGPGKDHETLRRRIENGAKELMFLPQS 132
Db 378 ATGEVMAIGRTYEESELKAIKRSLEYGVNHLGLPNCESFDLDYIKERISHODDERLFFIGE 437
QY 133 ELKRLKNEGNELOHNADEF-----LSDGHNERSIMTDLUYLSQYDGDAGDMKEKAK 185
Db 438 AIRKRTTLEELHNMTOIDYFLHKQNTIDIEHOLKEHOGDELYKAKVDG-FSDKTTA 496
QY 186 DLTELVRQRTIYLVONPKDCSAKAKLVNINCKGCGCOLHNHVYCFMIAVGTQRTLAES 245
Db 497 HRFNTEEEVYQLRBNNDIKPYKKV----- 522
QY 246 HNMRYATGCMETVRRPVSECTDRSGSGTGHMSGEVCKDNQOV---ELPIVDSVHPPR 301
Db 523 -----DTCAAEFESSTPYGYGTETENESIYDKELIIVLGSG--- 560
QY 302 PYLPLAVPEDLADRLVRYHGDPAVMWVSO--FVKYLLRPP-----WLEKEI 346
Db 561 ---PRIGOGVEFYATVH---AVMAIQAGYEALITVNNNPSTVDSISDKLFFELPT 614
QY 347 EEAATKLGFKRPVIGVNRRTDKVGAFAHPIEEYTVH-----VEEDF 390
Db 615 EEDVANNIINLEKPKGVV---QFGGQTAIINLADKLAKHGVKILGTSLENLNARADREF 670
QY 391 QLLARRMOYDKKRYVLAD-DEALLKEAKTKL-----PSYEF-----ISDMSWSAGL 438
Db 671 EALLRKINVPQOGKSATSPREALNAAEIGYVVVYPSYVYLGGRAMETVDN---DKEL 726

QY 439 HNRYTE-----NSLRGVILDIHLSQADFLVCTFSQCVRAVEIMQALHPDASANFRSLD 494
Db 727 ENVTQAKKASEHNVLD-RITKKELEVDAICGEEVITITGIMEHIER----- 775
QY 495 DIYFYGPNANQAIYHPORPTEGE-----IPMEPG-DITGVAGNHW---DG-YP 540
Db 776 ---AGVHSGSIANVPPQTLTEDELATLEDYTIKLAKGLMIGLINIQFIADGVYV 830
QY 541 KGVNRKLGRT 550
Db 831 LEVNPRESSRT 840

RESULT 5
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96796
R:Rheologos, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: D96796
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1871 <STO>
A:Cross-references: GB:AE005173; NID:96143906; PIDN:AF04452.1; GSPDB:GN00141
A:Gene: F28016.15
A:Map position: 1

Query Match 4.0%; Score 123; DB 2; Length 1871;
Best Local Similarity 25.0%; Pred. No. 4.7;
Matches 47; Conservative 40; Mismatches 73; Indels 28; Gaps 9;
QY 24 YIGHLVNDNHSNSSLK-ILAKLERLKOENEDLRMMAGSIRIPGPIDGSPASGR 82
Db 1065 HTGG-----EDNDHKKEEQKENVIAKAE-LNTEEDSFKKVEIEKODHGLKRSVQAK 1118
QY 83 VAALEE---OPMAKKEQIENYKQTKNG-----PGKHEI---LRRRIENGAKELMFF 129
Db 1119 RQTEBEKQTRAMEKNEYER-RKQTKDSLSKLAKGERPELGCHGRKCEDERIEL--- 1174
QY 130 LQSEL---KKIKNEGNELOHNADEFSLDGNHRSIMTDLUYLSQYDGDAGDMKEKAK 185
Db 1175 VETETSDHKEKKKDEYILRSQDTGKVDLGERRRRSKORKIKHSVEDEIGDQDEDAE 1234
QY 186 DLTELVR 193
Db 1235 EAAAVVSR 1242

RESULT 6
S22695
restin - human
C:Species: Homo sapiens (man)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 05-Nov-1999
C:Accession: S22695; S19853
R:Billbe, G.; Delabie, J.; Brueggen, J.; Richener, H.; Asseibergs, F.A.M.; Cerletti, N
EMBO J. 11, 2103-2113, 1992
A:Title: Restin: a novel intermediate filament-associated protein highly expressed in
A:Reference number: S22695; MUID:92289675
A:Accession: S22695
A:Molecule type: mRNA
A:Residues: 1-1427 <BIL>

A:Molecule type: DNA
 A:Residues: 1-376, 'V', 378-390, 'V', 392-576, 'L', 578-680, 'I', 682-1938 <DB>
 A:Cross-references: EMBL:X08065; NID:96785; PIDN:CAA30854.1; PID:96786
 R.Karrn, J.; Brenner, S.; Barnett, L.
 Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983
 A:Title: Protein structural domains in the *Caenorhabditis elegans* unc-54 myosin heavy chain
 A:Reference number: A93958; MUID:83273600
 A:Accession: A02993
 A:Molecule type: DNA
 A:Residues: 24-93, 'E', 95-97, 'R', 99-376, 'V', 378-388, 'GDV', 392-407, 'N', 409-473, 'G', 475-576
 C:Genetics:
 A:Gene: myo-1; CESP:R06C7.10
 A:Map position: 1
 A:Intons: 23/3; 114/3; 229/1; 264/1; 320/1; 857/3; 1745/3; 1814/1; 1892/3
 C:Superfamily: myosin heavy chain: myosin motor domain homology
 C:Keywords: actin binding; ATP: coiled coil; hydrolyase; methylated amino acid; muscle co
 F:87-773/Domain: myosin motor domain homology <MOT>
 F:177-184/Region: nucleotide-binding motif A (P-loop)
 F:660-682/Region: actin binding #status predicted
 F:764-778/Region: actin binding #status predicted
 F:846-1938/Domain: coiled coil #status predicted <COI>
 F:846-1160/Region: S2
 F:1161-1938/Region: light meromyosin
 F:128/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
 F:183/Binding site: ATP (Lys) #status predicted
 F:700,710/Active site: Cys #status predicted

Query Match 3.8%; Score 117; DB 1; Length 1938;
 Best Local Similarity 20.5%; Pred. No. 13; Mismatches 139; Indels 168; Gaps 21;

Matches 97; Conservative 70; Mismatches 139; Indels 168; Gaps 21;

36 SDHSSRELKTIKLERLKQONEDLR---MAGSLRIEGRIDGPAAGR-----V 83
 Db 1481 SRTSTVEFKLRSMNLSQIETLRREKKIFSQELRDINEQITG---GRYQVEHKSV 1537
 QY 84 RALEDFPMKAKEDIEYKQTKNG-PGKHDEILRRRIENGAKELMFLOSELKKNLBS 142
 Db 1538 RLLEQF---KDELQHALDEAEALAEESKVLK-----LQIEVOQIRSEIE 1580
 QY 143 NELQRADEFLSDGHNHR---SIMPDLVYLSOTDAGDMREKADLTDELVORRTIYQ 199
 Db 1581 KRIOKEEDEFENRRKKNRLESIOASL-----LEAKSKAELEAK----- 1622
 QY 200 NPKDSKAKKLYCNKGGCGGOLHNHVVYCFMAYGTQRTALLESNNRY--ATGSMET 257
 Db 1623 -----KKLETDIN-----QELIALDNANKANVDAQKLNKK 1652
 QY 258 VRRPVSETCTDRSGSTGHSCEVKKKNVQVLEPLIVDSVHRRPYL-----PLAVP 309
 Db 1653 LFPQVREL-----OGQVDEQRRREI---RENYLAEKRLALALLES 1692
 QY 310 EDLADRL-----VRHGDPAVMVVSQFVKYLIRPQMLEKEIEE 348
 Db 1693 EDLAHIEASDKKKQLEIEQALAKSSNTELLGNNA-----LSMKKRVENEVOQ 1743
 QY 349 AFKKIGFGRHVIGVHVRDYGAAE---AEHPLEETVHVE---EDFQLARRMAY- 399
 Db 1744 ARNELDEYINELKASERARKAAADRLAEERYOEENHNVDRORSLELNAKELQAK 1803
 QY 400 --DKRNVLTATDPALEKATKYPSYEFLSDNSISMSAGLN--RYTENSIG 449
 Db 1804 IDDAERAMIOFGAKALAK-----VEDRVSLEALHSEORRHOESIG 1846

RESULT 11
 T32568
 hypothetical protein F37C4.6 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T32568
 R.Langston, Y.; Sansone, J.; Wohldmann, P.
 submitted to the EMBL Data Library, December 1997
 A:Description: The sequence of C. elegans cosmid F37C4.

A:Reference number: Z21193
 A:Accession: T32568
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-544 <LAN>
 A:Cross-references: EMBL:AF036705; PIDN:AA95172.1; GSPDB:GN00022; CESP:F37C4.6
 A:Experimental source: strain Bristol N2; clone F37C4
 C:Genetics:
 A:Gene: CESP:F37C4.6
 A:Map position: 4
 A:Intons: 29/3; 230/2; 397/2; 484/3

Query Match 3.7%; Score 116; DB 2; Length 544;
 Best Local Similarity 19.6%; Pred. No. 2.7;

Matches 110; Conservative 72; Mismatches 213; Indels 166; Gaps 28;

25 IGHILVRDNDHSHSRELKTIKLERLKQONEDLRRAAGSLRIEGRIDGPAAGR 84
 Db 49 VGGAAVTEELVPGFRSRASYLSLIRPVVMOELNKKGLRYHT-RNPNSETP---IR 103
 QY 85 ALEDFPMKAKEDIEYKQTK-----NGPKDEILRRRIENGAKELMFLOSELKRL 137
 Db 104 NTHESLLGMDMAENKELAKSQRDAGNYPKENFISE-----IVHSFEQL 150
 QY 138 KNLGELQRAHDEFLSDGHNHRSIMTDLVYLSOTDAGDMREKAKDLTELVORRTIY 197
 Db 151 MDYERPLDQKPI-----HKLLPHLYLLFKT--VQPLGLRNAVDVFEELMTAPISK 197
 QY 198 LQN---PKDSKAK-----KLVCNINKGCGGOLHNHVVYCFMAYGTQRTALLES 246
 Db 198 INMKWESVLAATLCTDGVIGLAASPMDPGTGY-VLLNHVIG-----GLDEH 244
 QY 247 N--WRATGGMETVRRPVSETCTDRSGSTGHSCEVKKKNVQVLEP--IVDSVHRRP 302
 Db 245 KGAAGVYGGMGAVSNAIE-CAKSHGAEI-----YTQDQVQVLELDGNAVAKGV- 293
 QY 303 YLPLAVEDLADLVHVRDPAVMVVSQFVKYLIRPQMLEKEIEEATKLGKRPVIGV 362
 Db 294 --LSNGKELHSHVSNATPHVTF-----NHLVKES-LPEEFHNINQIDVTSVTKI 344
 QY 363 HVARTD-----KGAERAPRIEETVHVE-EDFQL-----ARR----- 396
 Db 345 NVAVKELPNEFLAKPNQGSSEPM--RHNOTTIHNCSEMOVNAVMDYKNGRYSRRPIEM 402
 QY 397 ---MOYDKKRVYLATDPALEKATKYPSYEFLSDNSISMSAGLNRYTENSIG 453
 Db 403 TTPSSVDRTIVDSADGHVLL-----FTQYTFPSKDGEMTEKKTEYAK----- 447
 QY 454 IHFLSQADFLVCTESSQVCRVAYEIMQALHPDASANF-----RSLDDIY----- 497
 Db 448 -HVFSEIDAVAPNFSSSV--IGVDILTP--PDIONFGITGNIFFHGSMSLDOLVYSRPI 502
 QY 498 -----YFGGPNAN 505
 Db 503 SKWSNYSFPLESLYLCGSGAH 523

RESULT 12
 I48176
 synaptonemal complex protein - golden hamster (fragment)
 C:Species: *Mesocricetus auratus* (golden hamster)
 C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 05-Nov-1999
 C:Accession: I48176
 J.Dobson, M.J.; Pearlmann, R.E.; Karlskaks, A.; Spyropoulos, B.; Moens, P.B.
 J. Cell Sci. 107, 2749-2760, 1994
 A:Title: Synaptonemal complex proteins: occurrence, epitope mapping and chromosome di
 A:Reference number: I48176; MUID:95181577
 A:Accession: I48176
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-845 <RES>
 A:Cross-references: GB:I32978; NID:9598124; PIDN:AA42039.1; PID:9598125
 C:Genetics:

Oy 317 VAVHDDPAVWVWQCVKXKLIRQPMLEKEIEIEATKKGKGFHPVGVHVRRT-----DKV 370
 Db 569 LRIEKELEHKLINLKRYKRLDRQ-----KQKEPAQK-----HKAQTELELLKRI 615
 Oy 371 GAFAF-----HPIEETVHVEDFOLLARMQ-VDK 402
 Db 616 REKSLRVKEFKELRYVERLEDEEELKEINVINKKLOIEEK 658
 RESULT 15
 F39845
 carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5), pyrimidine-repressibl
 N:Alternate names: carbamoyl-phosphate synthetase (catalytic chain) pyTAB
 C:Species: *Bacillus subtilis*
 C:Date: 20-Mar-1992 #sequence_revision 20-Mar-1992 #text_change 02-Sep-2000
 C:Accession: F39845; A69686
 R:Quinn, C.L.; Stephenson, B.T.; Switzer, R.L.
 J. Biol. Chem. 266, 9113-9127, 1991
 A>Title: Functional organization and nucleotide sequence of the *Bacillus subtilis* pyrmi
 A:Reference number: A59845; MUID:91225016
 A:Accession: F39845
 A:Molecule type: DNA
 A:Residues: 1-1071 <KUN>
 A:Cross-references: GB:M59757; NID:94887706; PIDN:AAA1270.1; PID:Q413390
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Acevedo, V.; Berte
 C.: Bron, S.; Boudillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.; Ch
 A.: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Laidinois
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
 Y. M.; Ogawa, K.; Ogilawa, A.; Oudega, B.; Park, S.H.; Parro, M.; Pohl, T.M.; Portelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeder, R.; Scofield, F.; Sekiguchi, J.; Sekowska, A.; Serot
 Akenuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A>Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044003
 A:Accession: A69686
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1071 <KUN>
 A:Cross-references: GB:Z59112; GB:AL009126; NID:92633902; PIDN:CAB13426.1; PID:92633925
 A:Experimental source: strain 168
 C:Genetics:
 A:Gene: pyTAB
 C:Function:
 A:Pathway: glutamate metabolism; pyrimidine nucleotide biosynthesis
 C:Family: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin
 C:Keywords: ligase; pyrimidine nucleotide biosynthesis
 F:9-1047/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homolo
 F:9-667/Domain: biotin carboxylase homology <BC1>
 F:553-985/Domain: biotin carboxylase homology <BC2>
 Query Match 3.6%; Score 111.5; DB 2; Length 1071;
 Best Local Similarity 19.0%; Pred. NO. 14;
 Matches 119; Conservative 83; Mismatches 204; Indels 219; Gaps 29;
 Oy 43 LSKILAKLERLKQONEDLRMAAGSL-----RIECPID-----GCP 78
 Db 318 IAKLAKIVAGSLDEEMNPVTKTYAAFEPAIDVVSIRPMPDFESANRKLGTOMK 377
 Oy 79 ASGRV-----RALEDQFMKAKOEIEN--YKQKKNRPGKADHLLRRRIENGAKELMFFLOS 132
 Db 378 ATGEVWATGRTLEESLLKAVRSLEADYVHLELKADADISDELLERIKKAGDERLFYLA 437
 Oy 133 ETKKLKLNLEGNELQHADEFSLDGHHERSIWTDLYLVSQTFG-----AGD--- 178
 Db 438 ATRRGVYAE-----DL--HEFSAL-DVFPFLHKLFGIYQFKEKLKANNAGTDIV 481
 Oy 179 -----WREKAKDTELVORRI-----TYL 198

```

Db      482 LRRAKELCFSDQIYSREKKMKKESELYSLRKQAGIAPVFKMIDTCALFESEETPEYSTYE 5411
OY      199 QNPKDCSKAKKRLVCYNINK--CSGYQGLHHVVYCFMIAVGTORTLLAESHNNRATGCM 255
OY      542 EENESVYTDKKSVMYWLGSGRILRIQGVFEFD-----YATVHSV-----MAIKAGY 5866
OY      236 ETVF--RPVSETCTDRSGSSSTGHSNGEYKDKNOVVELPLTVDSYHNPRLPLAV----- 3088
OY      587 EALIVNNNPETVSDFD-----ISDK-LYFEEPLTEDVHHIIDLBOQPMVVVQFC 6353
Db      309 -----PEDLADRLVYRHGDPRVMMVWSQFVKYLIRQPMLEKIEEATKKLGFKNPV-- 3539
Db      636 GQTAINLDELASRKYKLLGTS-----LEDIDRAED--RDKFEQALDELGVPOPLGK 6655
OY      360 IGVNHRPTDKGAEAFHPIDEEYTVHVEDEPQLLARRNOYDKKRVUJLATDOPALLFAKT 419
Db      666 TATSVNQNVAJSASDGG-----YPLVLRPSYVLGKRAEI---VYHEEELIIMKMAVK 735
OY      420 KYPSEYFISDSISWAGLHNRRTYENSLRGVILDIHFLSQADFLVCFPSSOVCAVAYEIM 479
Db      736 INPQHPLVD-----RY-----LTGKEIEVDVASDGEIVVP-----GIM 7700
OY      480 QALHMDASANFRSLDDIYFGGPNHNOIALYHPQRPTEG-----EIPMEBG-DIIG 5300
Db      771 EHIER-----AGVHSGDSIAVYPPQSLTEDIKKIKIEQYITALAKGLNTVG 8155
OY      531 VAG-----NHMDGYPKGVNKKLGGT 550
Db      816 LNIQIFVLDSGEVYVLEVNPPSSRT 840

```

Search completed: October 31, 2002, 13:09:03
Job time : 32 secs

179 -----WREKEAKDTELVRRI-----TYL 198

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 1, 2002, 21:34:02 ; Search time 2009 Seconds
(without alignments)
13930.215 Million cell updates/sec

Title: US-09-839-136-1

Perfect score: 1728
Sequence: 1 atgcgcacatgactgcttc.....atcccgagctcagaatga 1728

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlun:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_esthc:*
9: gb_estl:*
10: gb_estl2:*
11: gb_esthc:*
12: gb_estl3:*
13: gb_estl4:*
14: gb_estl5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	795	46.0	1066	14	BM928521 AGENCOURT
2	772.4	44.7	937	14	BO892101 AGENCOURT
3	753.6	43.6	940	14	BO653329 AGENCOURT
4	741.6	42.9	886	14	BO423860 AGENCOURT
5	718.4	41.6	752	14	BO604588 MI-P-CP1-AU124128
6	691.6	40.0	834	9	AU124128 AU124128

7	619.6	35.9	965	13	BI546364
8	573	33.2	936	12	BE893107
9	567.6	32.8	769	12	BE501191
10	557.8	32.3	702	13	BE973321
11	557.2	32.2	715	12	BE771085
12	555.8	32.2	656	14	BO552954
13	552.4	32.0	900	14	BO731540
14	536.8	31.1	653	12	BE619839
15	513.2	29.7	623	14	BO305885
16	512.2	29.6	1273	13	BO720741
17	508	29.4	1002	13	BI517671
18	502.6	29.1	621	14	BO305594
19	496.4	28.7	579	13	BI964023
20	495.6	28.7	600	14	BO305888
21	492.8	28.5	635	14	BO307749
22	489.4	28.3	1009	14	BE223892
23	487.2	28.2	573	9	A1122135
24	484.2	28.0	584	14	BO305887
25	483.2	28.0	888	12	BE731842
26	482.2	27.9	607	14	BO307741
27	475	27.5	594	14	BO308365
28	469.6	27.2	732	10	AW635821
29	463.6	26.8	970	14	BO926824
30	462.4	26.8	555	14	BO309171
31	460.2	26.6	529	10	BE543978
32	455.2	26.3	563	14	BO307126
33	455	26.3	598	14	BO305867
34	451.4	26.1	538	14	BO305873
35	449.8	26.0	534	14	BO307345
36	448.2	25.9	534	10	BE690071
37	445.4	25.8	697	9	AJ398890
38	436.2	25.2	505	14	BM717378
39	420	24.3	511	14	BO306837
40	419.2	24.3	826	12	BE246161
41	419	24.2	631	14	BO309231
42	414.8	24.0	484	10	AW387766
43	414.4	24.0	499	14	BO305595
44	412.4	23.9	587	14	BO305859
45	412.2	23.9	458	9	AA057866

ALIGNMENTS

BI546364	603188856
BE893107	601437002
BE501191	602546241
BE973321	602842563
BE771085	602719363
BO552954	H4019B10-
BO731540	AGENCOURT
BE619839	602618014
BO305885	MRO-BT200
BO720741	AGENCOURT
BI517671	603042284
BO305594	MRO-BT200
BI964023	1e63h10.y
BO305888	MRO-BT450
BO307749	MRO-BT450
BE223892	AGENCOURT
A1122135	u015d10.f
BO305887	MRO-BT200
BE731842	601568702
BO307741	MRO-BT450
BO308365	MRO-BT450
AW635821	b138903.w
BO926824	AGENCOURT
BO309171	MRO-BT450
BE543978	601070496
BO307126	MRO-BT450
BO305867	MRO-BT200
BO305873	MRO-BT200
BO307345	MRO-BT450
BE690071	uw64a03.y
AJ398890	AJ398890
BM717378	UI-E-E10-
BO306837	MRO-BT450
BE246161	601854133
BO309231	MRO-BT450
AW387766	MR4-ST011
BO305595	MRO-BT200
BO305859	MRO-BT200
AA057866	zk79c01.f

RESULT 1
BM928521
LOCUS
DEFINITION
AGENCOURT_6715254 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5798262
5', mRNA sequence.
ACCESSION
BM928521
VERSION
BM928521.1 (GI:19378900
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1066)
NIH-MGC <http://mgs.mcl.nih.gov/>.
AUTHORS
National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE
Unpublished (1999)
JOURNAL
Contact: Robert Strausberg, Ph.D.
COMMENT
Email: cgapbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM2025 row: p column: 07
High quality sequence stop: 666.
Location/Qualifiers
1. 1066

FEATURES
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5798262"
/clone.lib="NIH_MGC.100"
/issue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: liver; Vector: pORF7; Site_1: xhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally cloned
into EcoRI/xhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT      315 a      235 c      248 g      264 t      4 others
ORIGIN

Query Match      46.0%; Score 795; DB 14; Length 1066;
Best Local Similarity 89.8%; Pred. No. 5.6e-188;
Matches 885; Conservative 0; Mismatches 98; Indels 3; Gaps 3;

QY 366 AGCTAAGAGCTCTGCTTTTCTACAAAGTTCAGAAATTAAGAAATTTAGAGG 425
    |||||||
Db 1 AGCTAAGAGCTCTGCTTTTCTACAGAGTGAATTAAGAAATTTAGAGG 60

QY 426 AATGAACTCCAAGACATGACATGAATTTCTATCAGATTGGACATGAAAGTGC 485
    |||||||
Db 61 AATGAACTCCAAGACATGACATGAATTTCTTTGGATTAGACATCATGAAAGTGC 120

QY 486 TATTAATGAGGATCTATCTACTCTAGTCAAGATGGGAGGTGATGGCGTGA 545
    |||||||
Db 121 TATTAATGAGGATCTATCTACTCTAGTCAAGATGGGAGGTGATGGCGGGA 180

QY 546 GAGGGCCAAAGATCTGACAGAGTGTGCCAGGCGAATAACATATCTTACAGATCCCA 605
    |||||||
Db 181 AGAGGCCAAAGATCTGACAGAGTGTGCCAGGCGAATAACATATCTTACAGATCCCA 240

QY 606 GGAATGAGGCGAAGGAGGAGTGTGATTAATCAACAAGGCGTGGTATGGCTG 665
    |||||||
Db 241 GGAATGAGGCGAAGGAGGAGTGTGATTAATCAACAAGGCGTGGTATGGCTG 300

QY 666 TCAGCTTCATCATGTAGTGTACTGCTTTATGATTCATATGCAACCCAGGAACTGCG 725
    |||||||
Db 301 TCAGCTTCATCATGTAGTGTACTGCTTTATGATTCATATGCAACCCAGGAACTGCG 360

QY 726 CTTGGAATCTCACAATTTGGGCTAGCTGAGGAGTGGGAAAGCTGTTTACAGCTGT 785
    |||||||
Db 361 CTTGGAATCTCACAATTTGGGCTAGCTGAGGAGTGGGAAAGCTGTTTACAGCTGT 420

QY 786 AAGTGAGACGTGACAGACAGATCTGGCAGCTCCACTGACATTTGGTCAAGTAA 845
    |||||||
Db 421 AAGTGAGACGTGACAGACAGATCTGGCAGCTCCACTGACATTTGGTCAAGTAA 480

QY 846 GGAACAAAATGTTTCAGGTGTGAGCTCCCATTTGAGACAGTGTTCCTGCTCTCC 905
    |||||||
Db 481 GGAACAAAATGTTTCAGGTGTGAGCTCCCATTTGAGACAGTGTTCCTGCTCTCC 540

QY 906 ATATTATACCCCTGGCTGTCCCAAGACCTTGACAGATCCACTGTCAGATGCTGTGA 965
    |||||||
Db 541 ATATTATACCCCTGGCTGTCCCAAGACCTTGACAGATCCACTGTCAGATGCTGTGA 600

QY 966 TCCGTGAGTGTGGGTATCCAGATTGTTCAGATTAATTCGGCCCAACCCGTGCT 1025
    |||||||
Db 601 CCGTGCAGTGTGGGTATCCAGATTGTTCAGATTAATTCGGCCCAACCCGTGCT 660

QY 1026 GGAAGAAAGAAATGAGAGAGGCGACAAAGAGCTAGGCTTCAAAATTCAGATTATGGAGT 1085
    |||||||
Db 661 AGAAAGAAAGAAATGAGAGAGGCGACAAAGAGCTAGGCTTCAAAATTCAGATTATGGAGT 720

QY 1086 CCAATGTTAGACGACAGCAAAAGTGGGAGGAGGAGGAGCTTCCATTCCTTTAGAGAAATA 1145
    |||||||
Db 721 CCAATGTTAGACGACAGCAAAAGTGGGAGGAGGAGGAGCTTCCATTCCTTTAGAGAAATA 780
```

```
QY 1146 CAGGTCGACGTTGAAGAAAGACTTTCAGCTTCTTGCTCGCAGAAATGCAATGAATAAAA 1205
    |||||||
Db 781 CATGGTCAGATGTTGAGAAAGACTTTCAGCTTCTTGCTCGCAGAAATGCAATGAATAAAA 840

QY 1206 AAGGTCGATATTTGGCCACAGATGACCCCTGTTGTTAAAGAGCAAAAACAAAGTACCC 1265
    |||||||
Db 841 AAGGTCGATATTTGGCCACAGATGACCCCTGTTGTTAAAGAGCAAAAACAAAGTACCC 900

QY 1266 CAGTTTGAATTT-ATTAGATTAATCTATCT-CTTGGTCAGCTGGACTACATTAATCG- 1322
    |||||||
Db 901 CAGTTTGAATTTAATTAATGATTAATCTATCTTCCCTGGTCAGCGGACTTGCCTATCGN 960

QY 1323 ATATACAGAAATTCACCTCGGGGTG 1348
    |||||||
Db 961 ATATACCGGAAATTCCTCCCTCGGGG 986

RESULT 2
BQ892101      937 bp      mRNA      linear      EST 16-AUG-2002
LOCUS
DEFINITION
AGENCOURT 8743832 NIH_MGC_129 Mus musculus cDNA clone IMAGE:5389725
5', mRNA sequence.
ACCESSION
BQ892101
VERSION
BQ892101.1 GI:22284115
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 937)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Susan L. Sullivan, PhD.
cDNA Library Preparation: Resgen, Invitrogen Corp.
DNA sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1M413875 row: d column: 14
High quality sequence start: 130
High quality sequence stop: 780.
Location/Qualifiers
1..937
source
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5389725"
/clone.lib="NIH_MGC.129"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: olfactory epithelium; Vector:
pCMV-SPORT6.1.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Constructed by Resgen, Invitrogen Corp. Note: This
is a NIH_MGC Library."

BASE COUNT      268 a      215 c      209 g      244 t      1 others
ORIGIN

Query Match      44.7%; Score 772.4; DB 14; Length 937;
Best Local Similarity 90.0%; Pred. No. 2.4e-182;
Matches 838; Conservative 0; Mismatches 92; Indels 1; Gaps 1;

QY 736 CACAATTTGGCGCTAGCTACTGTTGGGATGGAAGCTGTGTTAGACCTGTAAGTGAAGC 795
    |||||||
Db 1 CACAATTTGGCGCTAGCTACTGTTGGGATGGAAGCTGTGTTAGACCTGTAAGTGAAGC 60

QY 796 TGCAAGACAGATCTGCGACCTGACATTTGGTCAGGTGAAGTGAAGCAAAAT 855
    |||||||
Db 61 TGCAAGACAGATCTGCGCTTCACCTGACACTGTGTCAGGTGAAGTGAAGCAAAAT 120
```


QY	886	GTTCGGGTGGTGGACCTCCCATCTGTACAGAGCTGTTCATCTCCGCTCCCATATTACC	915
Db	121	ATTCAAGTGGTGGACCTCCCATCTGTACAGAGCTGTTCATCTCCGCTCCCATATTACC	180
QY	916	CTGGCTGTCCCAAGAGACCTTGCACATGACCTTGTACAGATCATGATGATCCTGACAGT	975
Db	181	CTGGCTGTTCACAGAGACCTTGCACAGACCTTGTACAGATCATGATGATCCTGACAGT	240
QY	976	TGCTGGGTATCCCACTTTTGTCTCAAGTACTGTGATTCGCCCAACAACCTCGCTGGAAAGGA	1035
Db	241	TGCTGGGTATCCCACTTTTGTCTCAAGTACTGTGATTCGCCCAACAACCTCGCTGGAAAGGA	300
QY	1036	ATAGAAGAGCCCAACAAGAGCTAGGCTTTCACAACTTCCAGTTATTGAGTGCATGTTAGA	1095
Db	301	ATAGAAGAGCCCAACAAGAGCTAGGCTTTCACAACTTCCAGTTATTGAGTGCATGTTAGA	360
QY	1096	CGCACAGCAAAAGTGGAGCGGAGACGACACCTTCCATCCATGAGAGTAATACAGGCTGAC	1155
Db	361	CGCACAGCAAAAGTGGAGCGGAGACGACACCTTCCATCCATGAGAGTAATACAGGCTGAC	420
QY	1156	GTTGAAGAAGACTTTCACCTTCTTGCTGCGAGAAATGCAAGTGAGTAAAAAAGGCTGAT	1215
Db	421	GTTGAAGAAGACTTTCACCTTCTTGCTGCGAGAAATGCAAGTGAGTAAAAAAGGCTGAT	480
QY	1216	TTGGCCACACAGATGACCCCTGCTTTGTTTAAAGAGGCAAAACAAGTACCCAGTTATGAA	1275
Db	481	CTGGTACTGTATGATCCCTTCTTGTAAAGAGGCAAAACAAGTACTCCATTTATGAA	540
QY	1276	TTTATTACTGATTAATCTATCTCTTGCTGGACGTGAGTACATTAATGATATACAAAT	1335
Db	541	TTTATTACTGATTAATCTATCTCTTGCTGGACGTGAGTACATTAATGATATACAAAT	600
QY	1336	TCACCTTCGGGGTGTGATCCTGGATATACACTTTCCTCCAGGAGACTTCTCTAGTGT	1395
Db	601	TCACCTTCGGGGTGTGATCCTGGATATACACTTTCCTCCAGGAGACTTCTCTAGTGT	660
QY	1396	ACTTTTATCGCAGCTGTGTAGAGTGTCTTATGAAATCATGCAAGCCGTGATCCTGAT	1455
Db	661	ACTTTTATCGCAGCTGTGTAGAGTGTCTTATGAAATCATGCAAGCCGTGATCCTGAT	720
QY	1456	GCCCTTCGGAACTTCGCTCTTGTGATGACATCTATCTTGTGGAGGCCCAATGCCAC	1515
Db	721	GCCCTTCGGAACTTCGCTCTTGTGATGACATCTATCTTGTGGAGGCCCAATGCCAC	780
QY	1516	AACCAATATGCCATTTATTCCTCACCAACCTCGACATGAGAAGAAATCCCATGAGACT	1575
Db	781	AATCATGATTTGCTGTTATTCCTCACCAACCTCGACATGAGAAGAAATTCATGAGACT	840
QY	1576	GGAGATATTATTTGTTGTTGCTGGAATATCAT- GGGATGGCTATCTTAAAGTGTAAACAG	1634
Db	841	GGAGATATTATTTGTTGTTGCTGGAATATCATGGAAGTGGTGTATCTTAAAGTATCAACAG	900
QY	1635	AAAACCTGGGAAGAGAGGGCTATATCCCTCC	1665
Db	901	AAAACCTGGGAAGAGAGGGCTATATATCCCTTC	931
RESULT 3			
LOCUS	B0653329		
DEFINITION	B0653329	940 bp	mrna
ACCESSION	B0653329		Linear
VERSION	B0653329.1		EST 15-JUL-2002
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	NIH-MGC http://mgc.ncl.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		

COMMENT	CONTACT: Robert Strausberg, Ph.D. Email: c9apbs-remail.nih.gov Tissue Procurement: CGAP (Stanford) CDNA Library Preparation: Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://lmage.llnl.gov Plate: L16CM2445 row: k column: 03 High quality sequence stop: 700.
FEATURES	Location/Qualifiers
Source	1..940
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:6270458"
	/clone.lib="NIH.MGC.100"
	/tissue_type="hepatocellular carcinoma, cell line"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: Liver; Vector: pORF1; Site_1: XhoI; Site_2: EcoRI; CDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGAG(G). Size selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the Laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT	286 a 197 c 222 g 235 t
ORIGIN	
Query Match	43.6%; Score 753.6; DB 14; Length 940;
Best Local Similarity	91.1%; Pred. No. 1.2e-177;
Matches 856; Conservative	0; Mismatches 79; Indels 5; Gaps 5
QY	ACCTAAGACGCTCTGGTTTCTCTCAAAAGTGAAGTTGAAGAAATTTAGACG 425
DB	1 ACCTAAGACGCTCTGGTTTCTCTCAAGGTGAATGAAGAAATTTAGACG 60
QY	AAATGACGTCGCAAGACATGCAGATGAATTTCTATCAGATTTGGGACATGAAAGTC 485
DB	61 AAATGACGTCGCAAGACATGCAGATGAATTTCTTTGGATTAGACATGAAAGTC 120
QY	TATTAATGACGATCTATACTACCTCAGTCAACACAGATGGGCGAGTATGGCGTGA 545
DB	121 TATTAATGACGATCTATACTACCTCAGTCAACACAGATGGGCGAGTATGGCGG 180
QY	GGAGGCGAAGATCTGACAGAGCTGGTCCAGCGAGATTAACATATCTTCAGATCC 605
DB	181 AGAGGCGAAGATCTGACAGAGCTGGTCCAGCGAGATTAACATATCTTCAGATCC 240
QY	GGAGCGGCAAGGCAAGCAAGCAAGCTAGTGTATATCAACAAAGCGTGGCTAGGCG 665
DB	241 GGAGCGGCAAGGCAAGCAAGCTAGTGTATATCAACAAAGCGTGGCTAGGCG 300
QY	TCAGCTCATCTAGTGTAGTGTACTGCTTATGATTCGATATGGCACCCAGCAACTGC 725
DB	301 TCAGCTCATCTAGTGTAGTGTACTGCTTATGATTCGATATGGCACCCAGCAACTGC 360
QY	CTTGGATCTCAAAATTTGGCGCTAGCGCTAAGTGGATGGGAAACTGTGTTAGACCTG 785
DB	361 CTTGGATCTCAAAATTTGGCGCTAAGTGGATGGGAAACTGTGTTAGACCTG 420
QY	AAGTCAGACGTCACAGACAGATCTGGCGAGCTCCACTGACATTTGGTCAGGTCAACT 845
DB	421 AAGTCAGACGTCACAGACAGATCTGGCGAGCTCCACTGACATTTGGTCAGGTCAACT 480
QY	GGACAAAATGTTTCAAGTGTGAGCTCCCAATTTGATGACATGTTTCATCTCGTCTCC 905
DB	481 GGACAAAATGTTTCAAGTGTGAGCTCCCAATTTGATGACATGTTTCATCTCGTCTCC 540
QY	ATATTTACCCCTGGCTGTCCACAGAAAGCTTGACATGATCTTGTAGAGTCCATGCTGA 965

Db 541 ATATTTACCTTGGCTGTACACAGACCTTCGACATGCTTGTACGATGCTGCA 600
Qy 966 TCCGTGAGTGTGTGGTATCCGATTTTCAAGTACTGATTCGCCACACCTGGCT 1025
Db 601 CCGTGAAGTGTGGTGTCTCAGTTTGTCAATAGTTAGTCCGCCACAGCTTGGCT 660
Qy 1026 GGAAGAGAAATAGAGAGGCCACCAAGAGTACGCTTAAACATTCACGTTATGGAGT 1085
Db 661 AGAAAAAGAAATAGAGAGGCCACCAAGAGCTTGGCTTC- AACATCACTTATTTGGAGT 719
Qy 1086 CCAATGTAGACGACAGACAAAGTGGAGGCGAAGCGCTTCATCCATTCAGAGATA 1145
Db 720 CCAATGTAGACGACAGACAAAGTGGAGAGAGAGTCCCTTCATTCCTCAAGAGTA 779
Qy 1146 CAGCGT- GCAAGTGTAGAGAACTTTCAGCTTCTGCTCGACAGATGCAAGTATATAA 1204
Db 780 CAGTGTGAGTGTAGAGAACTTTCAGCTTCTGCTCGACCAAAATGCAAGTATGACCAA 839
Qy 1205 AAAGGCTGTA-TTTGGCCACAGATGACCTTCTTGT- TAAAGAGCAAAAAAGATA 1262
Db 840 AAAGAGTGTATTTGGCCACAGATGACCTTCTTATTTAAGAGCAAAAAAGATA 899
Qy 1263 CCCCAGTTATGAA-TTTATAGTATTAATCTATCTCTG 1301
Db 900 CCCCATTATGAATTTATATGATTAATCTATCTCTG 939

RESULT 4
BO423860 886 bp mRNA linear EST 23-MAY-2002
LOCUS BO423860
DEFINITION AGENCOURT_7894821 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6158200
5', mRNA sequence.

ACCESSION BO423860
VERSION BO423860.1 GI:21119175
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 886)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: ATCC/DCPD/DPF
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13505 row: e column: 17
High quality sequence stop: 641.

FEATURES
source
1.886
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6158200"
/clone_lib="NIH-MGC-72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Skin; Vector: pCMV-Sport6; Site_1: NCI; Site_2: Sall; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."

BASE COUNT 273 a 185 c 212 g 215 t 1 others
ORIGIN

Query Match 42.9%; Score 741.6; DB 14; Length 886;
Best local similarity 91.2%; Pred. No. 1.2e-174;
Matches 798; Conservative 0; Mismatches 75; Indels 2; Gaps 1;

Qy 408 ATTAAGAAATTTAGANGAATGAACTCCAAAGACATGACAGATGAATTTCTACATTT 467
Db 1 ATTAAGAAATTTAGANGAATGAAATGAACTCCAAAGACATGAGATGAATTTCTTTGGATT 60
Qy 468 GGGACATCATGAAAGGCTATATGACGGATCTATCTACCTAGTCAAAAGATGGGGC 527
Db 61 AGGACATCATGAAAGGCTATATGACGGATCTATCTACCTAGTCAAAAGATGGGGC 120
Qy 528 AGGTGATTTGGCGTGAAGAGAGGCCAAAGATCTGACAGAGCTGGTCCAGCGAGATAAC 587
Db 121 AGGTGATTTGGCGGGAAGAGAGGCCAAAGATCTGACAGAACTGGTTCACGCGAGATTAAC 180
Qy 588 ATATCTTCAGAAATCCCAAGAGCTGACAGCAAGCCAAAGAGTATGTGTATATACAA 617
Db 181 ATATCTTCAGAAATCCCAAGAGCTGACAGCAAGCCAAAGAGTGTGTATATTAACAA 240
Qy 648 AGGCTGGCTATGGCTGACAGCTCCATCATATGATGATGATGATGATGATGATGATG 707
Db 241 AGGCTGGCTATGGCTGACAGCTCCATCATATGATGATGATGATGATGATGATGATG 300
Qy 708 CACCCAGCAACACTGCTGGATCTGACAAATTCGCAATTTGGCGTACCTTACGCGGATGGA 767
Db 301 CACCCAGCAACACTGCTGATCTTGGAAATCTGAGATTTGGCGTATGCTACTGATGATGGA 360
Qy 768 AACTGTGTTAGACCTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 827
Db 361 GACTGTATTTAGGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 420
Qy 828 TTGCTGAGTGAAGTAAAGACCAAAATGTTCAAGTGGTTGAGCTCCCATTTAGACAG 887
Db 421 CTGCTGAGTGAAGTAAAGACCAAAATGTTCAAGTGGTTGAGCTCCCATTTAGACAG 480
Qy 888 TGTTCATCTGCTGCTCCATATTTTACCCCTGCTGTCGAGAAAGCTTGCAGATGACT 947
Db 481 TCTTCATCTGCTGCTCCATATTTTACCCCTGCTGTCGAGAAAGCTTGCAGATGACT 540
Qy 948 TGTTCATCTGCTGCTCCATATTTTACCCCTGCTGTCGAGAAAGCTTGCAGATGACT 1007
Db 541 TGTTCATCTGCTGCTCCATATTTTACCCCTGCTGTCGAGAAAGCTTGCAGATGACT 600
Qy 1008 TCCGCCAACAACCTGCTGGTGAAGAAATAGAGAGGCGCACCAAGACCTGCTTCAA 1067
Db 601 CCGCCCAACAACCTGCTGGTGAAGAAATAGAGAGGCGCACCAAGACCTGCTTCAA 660
Qy 1068 ACATTCAGTTATTTGAGTCCATGTTAGACGACAGCAAAAGTGGAGCGAAGCCTTT 1127
Db 661 ACATTCAGTTATTTGAGTCCATGTTAGACGACAGCAAAAGTGGAGCGAAGCCTTT 720
Qy 1128 CCAATCCATTTGAGTATACAGCGGTGACGCTGTAAGAAAGTTCACCTTCTGCTGCGAG 1187
Db 721 CCAATCCATTTGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 780
Qy 1188 AATGCAAGTGTATTAAGAAAGAGGTGATTTGGCCACAGATGAGCC--TTCCTTTGTATAA 1245
Db 781 AATGCAAGTGTATTAAGAAAGAGGTGATTTGGCCACAGATGAGCCCTTCTTTATTAAG 840
Qy 1246 GAGGCAAAAACAAAGTACCCCACTTATGAAATTTAT 1280
Db 841 GAGGCAAAAACAAAGTACCCCACTTATGAAATTTT 875

RESULT 5
BO604588 752 bp mRNA linear EST 24-JUN-2002
LOCUS BO604588
DEFINITION MI-P-CPI-nzc-f-18-0-UI-s1 MI-P-CPI Sus scrofa cDNA clone
MI-P-CPI-nzc-f-18-0-UI 3', mRNA sequence.

ACCESSION BO604588
VERSION BO604588.1 GI:21551314
KEYWORDS EST.
SOURCE Sus scrofa pig.
ORGANISM Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

1 (bases 1 to 752)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
Contact: Tugale, CX
97044477

Email: ckungj@iastate.edu
Tissue Procurement: Dr. Chris Tugule, Iowa State University
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 FORWARD
SOLYA-yes.

FEATURES	Location/Qualifiers
source	1. .752

`/lib.host="PH108 (Life Technologies)"` with a modified `/note=Vector: PT73D-Pac (Pharmacia)` with a modified polylinker. Site_1: Not I; Site_2: EcoRI. The MI-P-CPI library is normalized library derived from the MI-P-CPI library, ultimately derived from uterus tissue. For a detailed description of the library from which this clone was derived, please visit our web site at <http://pfigest.genome.iastate.edu/>. The procedure used to create this library has been previously described (Bonaldo Lennon and Soares, Genome Research 6: 791-806, 1996).

	TAG_SEQ-AGTCCATCG"			
BASE COUNT	221 a	154 c	189 g	187 t
ORIGIN				

Query Match	41.6%;	Score 718.4;	DB 14;	Length 752;
Best Local Similarity	98.4%;	Pred. NO. 7.4e-169;		
Matches 736; Conservative	0;	Mismatches 11;	Indels 1;	Gaps 1;

Oy 377 TCTGCGTTTTTTCTACAAAGTGAGTTGAAGAATTTAAAGCATTTTACAAGGAATGCACTCC 436
| | | | |
Db 5 TTATTATTTTATTTTGTCAAGCTGAGTTGAAGAATAAAGCAATTTACAAGGAATGCACTCC 64

Qy 437 AAGACATCGAGTGAATTTCTATCAGATT-TGGGACATCATGAAGGCTATTAATGACG 495

Db 65 AAGACATCGAGTGAATTTCTATCAGATTNTGGACATCATGAAGGCTATTAATGACG 124

Qy 496 GATCTATACCTACCTCAGTCAAAACAGATGGGGCAGCTGATTGGCGTCAAAAAGAGGCCAAA 555
 |||||
 Db 125 GATCTATACCTACCTCAGTCAAAACAGATGGGGCAGCTGATTGGCGTCAAAAAGAGGCCAAA 184

556 GATCTGACACAGCTGTGTCACCGGAGATTAACATATCTTCAGAACTCCAAAGGACTGCAGC 615
 |||
 185 GATCTGACACAGCTGTGTCACCGGAGATTAACATATCTTCAGAACTCCAAAGGACTGCAGC 244

245 AAGCCAGAAGCTAGTGTATATATCAACAAGGCTGTGGCTGTACAGCTCCAT 304

305 CAGTGTGTTACTGCTTATGATTGCATATGGCACCACCGAAGACTCATCTTGGAACTC 364

Figure 1

Db	365	CACAAATTGGCCCTACGCTACTGGGGGATGGAAACTGTGTGTTAAGCTGTAAAGTGACAGC	424
QY	796	TGCACAGACAGATCTGGCAGCTCCACTGGACATTTGGTCAGCTGGAATTAAGACAAAAT	855
Db	425	TGCACAGACAGATCTGGCAGCTCCACTGGACATTTGGTCAGCTGGAATTAAGACAAAAT	484
QY	856	GTTTCAGGTGCTTGAAGCTCCCAATTTGAGACAGTGTTCATCTCTGCTCCATATTTAACCC	915
Db	485	GTTTCAGGTGCTTGAAGCTCCCAATTTGAGACAGTGTTCATCTCTGCTCCATATTTAACCC	544
QY	916	CTGGCTGTCCAGAAAGACTTTGCAGATCGACTTTGACAGTTCATGTGGTGAATCTGCAGTG	975
Db	545	CTGGCTGTCCAGAAAGACTTTGCAGATCGACTTTGACAGTTCATGTGGTGAATCTGCAGTG	604
QY	976	TGGTGGGATATCCCAAGTTTGTCAAGTACTTGAATTTGGCCACAAACCCTGGCTGGAAAAGGAA	1035
Db	605	TGGTGGGATATCCCAAGTTTGTCAAGTACTTGAATTTGGCCACAAACCCTGGCTGGAAAAGGAA	664
QY	1036	ATAACAGAGGCGCACCAAGAACTAGGCTTCAAAATCCAGTTATTGAGTGCATGTTTGA	1095
Db	665	ATAACAGAGGCGCACCAAGAACTAGGCTTCAAAATCCAGTTATTGAGTGCATGTTTGA	724
QY	1096	CGCACAGCAAAAGTGGAGCGGGAAGCAG	1123
Db	725	CGCACAGCAAAAGTGGAGCGGGAAGCAG	752

RESULT 6	834 bp	mRNA	linear	EST 01-AUG-2002
AUI24128				
LOCUS	AUI24128	NT2RM2	Homo sapiens cDNA clone NT2RM2001714 5', mRNA	
DEFINITION	sequence.			
ACCESSION	AUI24128			
VERSION	AUI24128.1	GI:10948844		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			

REFERENCE
AUTHORS

1 (bases 1 to 834)
Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Salto, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and

TITLE
HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saiko,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S., Isogai,T.)

COMMENT
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute

Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hrl.co.jp

Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

```
source
1. 834
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cdate="NM726201714"
```

```

/cell_line="NT2"      from uninduced NT2 neuronal
/notes="Vector: NME1ASEL3: cRNA

```

BASE COUNT ORIGIN	precursor cells	others
249 a	187 c	181 g
		215 t

[illegible]

Best Local Similarity 90.7%; Pred. No. 3.8e-162;
Matches 747; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

```
OY 905 CATATTACCCTGGCTGCCAGAAAGACCTTGCAGATGCAGCTTGACAGTCCATGGTG 964
DB 1 CATATTACCCTGGCTGCCAGAAAGACCTTGCAGATGCAGCTTGACAGTCCATGGTG 60
OY 965 ATCTGTCAGTGTGGGTATCCAGATTGTGCAAGTACTGATGCGCCCAACAACCTGGC 1024
DB 61 ACCCTGACAGTGTGGNGGTGTCTCAAAATACTGATGCGCCCAACAACCTGGC 120
OY 1025 TGGAAAGAAATAGAGAGGCCACCAGAACCTAGCTTCAACATCCAGTTATGGAG 1084
DB 121 TGGAAAGAAATAGAGAGGCCACCAGAACCTAGCTTCAACATCCAGTTATGGAG 180
OY 1085 TCCATGTTGAGCGACAGCAAAAGTGGGAGGAGGACGACCTCCATCCATTTGAGGAT 1144
DB 181 TCCATGTTGAGCGACAGCAAAAGTGGGAGGAGGACGACCTCCATCCATTTGAGGAT 240
OY 1145 ACACGGTGCAGCTTGAAGAAGACTTTCACCTTCTGCTGCAGAAATGCAAGTGAATAA 1204
DB 241 ACATGTTGATGTGGAAGAATTTTACCTTCTGCTGCAGAAATGCAAGTGAATAA 300
OY 1205 AAGAGGTGATTTGGCCACAGATGACCTTCTTGTAAAGAGGCAAAACAAAGTACC 1264
DB 301 AAGAGGTGATTTGGCCACAGATGACCTTCTTGTAAAGAGGCAAAACAAAGTACC 360
OY 1265 CCAGTTATGATTTATGATGATCTATCTATCTGCTGACGTGACATACATTAATCAT 1324
DB 361 CCATTTATGATTTATGATGATCTATCTATCTGCTGACGTGACATACATTAATCAT 420
OY 1325 ATACAGAAATTCACCTTGGGGTGTGATCTGATATACACTTCTTCCAGGACGACT 1384
DB 421 ATACAGAAATTCACCTTGGGGTGTGATCTGATATACACTTCTTCCAGGACGACT 480
OY 1385 TCTTATGTTGATTTTATGATGAGCTGTGATGATCTTATGATATCAATCAAGCC 1444
DB 481 TCTTATGTTGATTTTATGATGAGCTGTGATGATCTTATGATATCAATCAAGCC 540
OY 1445 TGCATGCTGATGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1504
DB 541 TGCATGCTGATGCTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
OY 1505 CAATGCGCCAGCAACAATTCGATTTATCTCACCACCACTGGAAGTGAAGAAATCC 1564
DB 601 AGAATGCCCAATCAATTCGATTTATGATGATGATGATGATGATGATGATGATG 660
OY 1565 CCATGGAACCTGGAGATATTTATGTTGGTGGTGAATACAGTGGATGCTTCTAAG 1624
DB 661 CCATGGAACCTGGAGATATTTATGTTGGTGGTGAATACAGTGGATGCTTCTAAG 720
OY 1625 GTGTTAACGAAGAACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1684
DB 721 GTGTTAACGAAGAACTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 779
OY 1685 TAGAAACAGTCAAGTACCCACATATCCGAGGCTGACAAGTAA 1728
DB 780 TAGAAACAGTCAAGTACCCACATATCTGAGAGTGAAGAAATA 823
```

RESULT 7
B1546364 965 bp mRNA linear EST 05-SEP-2001
LOCUS 60318856F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5260284 5',
DEFINITION mRNA sequence.
ACCESSION B1546364
VERSION B1546364.1 GI:15433676
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 965)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palokvits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shihai
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
<http://image.lnl.gov>
Plate: LLM11655 row: p column: 13
High quality sequence stop: 828.

FEATURES
source location/Qualifiers
1..965
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5260284"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dt primed using primer 5'-gtttttttttttttttttt-3',
size-selected for average insert size 2.5 kb and
normalized for R0F 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NHRI, National
Institutes of Health). Note: this is a NIH-MGC Library."
BASE COUNT 298 a 213 c 207 g 247 t
ORIGIN

Query Match 35.9%; Score 619.6; DB 13; Length 965;
Best Local Similarity 91.2%; Pred. No. 4e-144;
Matches 680; Conservative 0; Mismatches 64; Indels 2; Gaps 2;

```
OY 973 GTGTGGTGGGTATCCAGTTGTCAAGTACTGATTCGCGCCCAACCTGGCGTGAAG 1032
DB 4 GTGTGGTGGGTATCCAGTTGTCAAGTACTGATTCGCGCCCAACCTGGCGTGAAG 63
OY 1033 GAAATAGAAAGAGCCACCAAGAAGCTAGGCTTCAACATCCAGTTATGAGTCCATGT 1092
DB 64 GAAATAGAAAGAGCCACCAAGAAGCTAGGCTTCAACATCCAGTTATGAGTCCATGT 123
OY 1093 AGAGCCACAGACAAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1152
DB 124 AGAGCCACAGACAAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 183
OY 1153 CAGGTTGAAGAACTTTCAGCTTCTGTCGACAGAAAGTGAAGTGAAGTGAAGTGA 1212
DB 184 CAGGTTGAAGAACTTTCAGCTTCTGTCGACAGAAAGTGAAGTGAAGTGAAGTGA 243
OY 1213 TATTTGGCCAGATGACCTGCTTTGTTAAAGAGGCAAAACAAATACCCAGTTAT 1272
DB 244 TATTTGGCCAGATGACCTGCTTTGTTAAAGAGGCAAAACAAATACCCAGTTAT 303
OY 1273 GAATTTATTTAGTATCTATCTTCTGTTGAGTGCAGTACATATGATATACAGAA 1332
DB 304 GAATTTATTTAGTATCTATCTTCTGTTGAGTGCAGTACATATGATATACAGAA 363
OY 1333 AATTCAGTTCGGGGTGTGATCTGATTTACACTTTCCTCCAGGACGAGCTTCTAGT 1392
DB 364 AATTCAGTTCGGGGTGTGATCTGATTTACACTTTCCTCCAGGACGAGCTTCTAGT 423
OY 1393 TGTACTTTTTCATCCAGGTGTGATGCTGTTATGAATCATGCAAGCCCTGCATCT 1452
DB 424 TGTACTTTTTCATCCAGGTGTGATGCTGTTATGAATCATGCAAGCCCTGCATCT 483
OY 1453 GATGCTCTGCGAAGCTTCCGTTCTTTGATGACATCTATTTTGGAGGCCCAATGCC 1512
DB 1512 GATGCTCTGCGAAGCTTCCGTTCTTTGATGACATCTATTTTGGAGGCCCAATGCC 1512
```

Db 484 GATGCTCTGCGAAATTCATCTTTAGATGACATCTACTATTTTGGGGCCAGAAATGCC 543
1513 CACAACCAATTCGCAATTTNCTCACCACCTGCAACTGAGAGAAATCCCATGGAA 1572
Db 544 CACATCAATATTCGCAATTTNCTCACCACCTGCAACTGAGAGAAATTCGCAATGGAA 603
Qy 1573 CCTGAGATATTTATTTGGTGTGCTGTAAGAACTACTGAGTGGCTATTCGTAAGGTGTTAAC 1632
Db 604 CCTGAGATATTTATTTGGTGTGCTGTAAGAACTACTGAGTGGCTATTCGTAAGGTGTTAAC 663
Qy 1633 AGAAATCGGAGAGAGCGGCTATTCCTCC-TACAAGTTCGAGAGAGATTA-GAAA 1650
Db 664 AGGAAATTCGAGAGAGCGGCTATTCCTCCCTTACAAGTTCGAGAGAGATTAAGAAC 723
Qy 1691 CAGTCAGTACCCCAATATCCCGAG 1716
Db 724 CGGTCAAGTACCCCAATATCCCGAG 749

RESULT 8
BE893107 936 bp mRNA linear EST 20-OCT-2000
LOCUS 601437002F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3922104 5',
DEFINITION mRNA sequence.
ACCESSION BE893107
VERSION BE893107.1 GI:10354133
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 936)
AUTHORS NIH-MGC http://mgs.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHAM9756 row: C column: 01
High quality sequence start: 3
High quality sequence stop: 696.
Location/Qualifiers
1. 936
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3922104"
/clone_lib="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-Sport6; Site_1: Not;
Site_2: salt; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 260 a 202 c 244 g 230 t
ORIGIN

Query Match 33.2%; Score 573; DB 12; Length 936;
Best Local Similarity 89.8%; Pred. No. 1.8e-132;
Matches 660; Conservative 0; Mismatches 70; Indels 5; Gaps 4;

Qy 351 GAGGATTCGAAATGAGCTCTGCTTTTCTACAAGTGAAGTGAAGAAATT 410
Db 8 GTGGCTGAAATGAGCTCTGCTTTTCTACAAGTGAAGTGAAGAAATT 67
Qy 411 AAGGATTCGAAATGAGCTCTGCTTTTCTACAAGTGAAGTGAAGAAATT 470
Db 68 AAGGATTCGAAATGAGCTCTGCTTTTCTACAAGTGAAGTGAAGAAATT 127

Qy 471 ACATCATGAAGGCTCTAATTAAGAGATCTTACTACTGATGCAACAGATGGGCGAG 530
Db 128 ACATCATGAAGGCTCTAATTAAGAGATCTTACTACTGATGCAACAGATGGGCGAG 187
Qy 531 TGATTCGCTGTAAGAGAGGCGCAAGATCTGACAGACTGGTTCAGAGGAGAAATACATA 590
Db 188 TGATTCGCTGTAAGAGAGGCGCAAGATCTGACAGACTGGTTCAGAGGAGAAATACATA 247
Qy 591 TCTTCAGATTCGCAAGAGATCTGACAGCAAGCCAGAGACTACTGTGTAATATCAACAAAG 650
Db 248 TCTTCAGATTCGCAAGAGATCTGACAGCAAGCCAGAGACTACTGTGTAATATCAACAAAG 307
Qy 651 CTGTGCTATGCTGTGACAGCTCCATCATGTAGTGTACTGCTTATGATATGATGCGAC 710
Db 308 CTGTGCTATGCTGTGACAGCTCCATCATGTGTGTCTGCTTATGATATGATGCGAC 367
Qy 711 CCAGCAACACTCGCTGTAAGTCAATTCAGATTCAGCTGAGGAGATGGGAAAC 770
Db 368 CCAGCAACACTCATCTGTAAGTCAATTCAGATTCAGCTGAGGAGATGGGAAAC 427
Qy 771 TGTGTTAGACTCTGTAAGTCAATTCAGATTCAGCTGAGGAGATGGGAAAC 830
Db 428 TGTATTTAGGCTGTGTAAGTCAATTCAGATTCAGCTGAGGAGATGGGAAAC 487
Qy 831 GTCAAGTGAAGTGAAGGACAAATATGTTACAGTGTGAGTGTGAGTGTGAGACAGTGT 890
Db 488 GTCAAGTGAAGTGAAGGACAAATATGTTACAGTGTGAGTGTGAGTGTGAGACAGTGT 547
Qy 891 TCATCTCTGCTCTCATATTTACCCCTGGCTGCTCCAG-AGAGCTTCGACATGACATTTG 949
Db 548 TCATCTCTGCTCTCATATTTACCCCTGGCTGCTCCAG-AGAGCTTCGACATGACATTTG 606
Qy 950 TACGATTCATG--TGATCTGCAAGTGTGCTGATTCGATTTGTCAGATTTGAT 1007
Db 607 TACGATTCATG--TGATCTGCAAGTGTGCTGATTTGTCAGATTTGATTTG 666
Qy 1008 TCGGCCCAACCCCGCTGGAAGAAATAGAGAGCGCAAGAGCTTACCTCAA 1067
Db 667 TCGGCCCAAC-ACCTGCTTGAAGAAATAGAGAGCGCAAGAGCTTACCTCAA 725
Qy 1068 ACATCCAGTTTATGG 1082
Db 726 ACATCCAGTTTATGGG 740

RESULT 9
BG501191 769 bp mRNA linear EST 27-MAR-2001
LOCUS 60254624F1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4668562 5',
DEFINITION mRNA sequence.
ACCESSION BG501191
VERSION BG501191.1 GI:13462708
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 769)
AUTHORS NIH-MGC http://mgs.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LHC1479 row: a column: 11
High quality sequence stop: 758.

FEATURES
source

Location/Qualifiers
1. 769
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4668562"
/clone_1lb="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
site:1: SfiI (99ccctcggcgc); site:2: SfiI (99ccctcggcgc)
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATCTGAGAGCCGCGGCGCATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
library."

BASE COUNT 240 a 164 c 166 g 199 t

ORIGIN

Query Match 32.8%; Score 567.6; DB 12; Length 769;
Best Local Similarity 87.0%; Pred. No. 3.8e-131;
Matches 658; Conservative 0; Mismatches 94; Indels 4; Gaps 3;

OY 975 GTGGTGGTATCCCATTTGTCTCAAGTACTGATTCGCCCAACCTGGCTGGAAAAGA 1034
|||||
DB 1 GTGGTGGTGTCTCAATTTGTCAATTAATCTGATCCGCCACGCTTGGCTAGAAAAGA 60
OY 1035 AATAGAAGAGCCACCAAGAGCTAGCTTCAAAATCCAGTTATTTGAGTCCATGTAG 1094
|||||
DB 61 AATAGAAGAGCCACCAAGAGCTTGTCAAAATCCAGTTATTTGAGTCCATGTAG 120
OY 1095 AGCCACAGACAAAGTGGAGCGGAGAGCTTCCATCCATTTGAGGAATACCGTGCA 1154
|||||
DB 121 ACCACAGACAAAGTGGAGAGAGAGCTTCCATCCATTTGAGGAATACCGTGCA 180
OY 1155 CGTTGAAGAGAGCTTCACTTCTCTGCGAGAAATGAGTGAATAAAAGGCTGA 1214
|||||
DB 181 TGTGAAGAGAGCTTCACTTCTCTGCGAGAAATGAGTGAATAAAAGGCTGA 240
OY 1215 TTTGGCCAGAGATGACCTGCTTTGTTAAAAGGCAAAAACAAGTACCCCGATTGTA 1274
|||||
DB 241 TTTGGCCAGAGATGACCTGCTTTGTTAAAAGGCAAAAACAAGTACCCCGATTGTA 300
OY 1275 AATTATTTAGTGAATCTATCTCTGTCAGTGCAGTCAATATGATATACAGAAA 1334
|||||
DB 301 AATTATTTAGTGAATCTATCTCTGTCAGTGCAGTCAATATGATATACAGAAA 360
OY 1335 TTTCATTTGGGGTGTGATCTGATATACACTTCTCTCCAGGAGACTTCTAGTGTG 1394
|||||
DB 361 TTTCATTTGGGGTGTGATCTGATATACACTTCTCTCTCAGGAGACTTCTAGTGTG 420
OY 1395 TACTTTTTCATGCGAGCTGTAGAGTTCCTTAATGAAATCATGCAAGCCGTGCATCTGA 1454
|||||
DB 421 TACTTTTTCATGCGAGCTGTGAGTGTGCTTAATGAAATCATGCAAGCCGTGCATCTGA 480
OY 1455 TGGCTTCGGAAGTTCGCTTTCTTGGATGACATCTACTATTTTGGAGGCCCAATGCC- 1513
|||||
DB 481 TGGCTTCGGAAGTTCGCTTTCTTGGATGACATCTACTATTTTGGAGGCCCAATGCCCA 540
OY 1514 ACAGCAAAATGCAATTTATTCCTCACCACCTGCAAGTGAAGAGAAATCCCATGGAAC 1573
|||||
DB 541 ACATCAAAATGCAATTTATTCCTCACCACCTGCAAGTGAAGAGAAATCCCATGGAAC 600
OY 1574 CTGAGATATATTTGGTGTGGCTGGAATACATGAGATGGCTATCTCT-AAAAGTGTAAAC 1632
|||||
DB 601 CTGAGATATATTTGGTGTGGCTGGAATACATGAGATGGCTATCTCTAAAAGTGTACAC 660
OY 1633 AGAAAATGGGAAGAGGAGCTATATCCCTCTACAAAAGTTGAGAGAGATAGAAACA 1692
|||||

DB 661 AGGAAATTTGGAGAGAGCGGCTATATCTCTAC-TACAAAGTCGAGACAGATAGAAACG 718
OY 1693 GTCAAGTACCCCATATATCCGAGCTGACAGTAA 1728
|||||
DB 719 GTCCAGTACCCCATATATCTAGGTGAGAAATAGTCA 754

RESULT 10
BG973221 702 bp mRNA linear EST 12-JUN-2001
LOCUS 602842563f1 NCI_CGAP_Mem4 Mus musculus cDNA clone IMAGE:4978383 5',
DEFINITION mRNA sequence.
ACCESSION BG973221.1 GI:14360858
VERSION BG973221.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 702)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furch
Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: LLAM10974 row: n column: 16
High quality sequence stop: 694.

FEATURES
source
Location/Qualifiers
1. 702
/organism="Mus musculus"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:4978383"
/clone_1lb="NCI_CGAP_Mem4"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: PCMV-SPORE6; Site:1: SalI;
Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furch.
NIH Reference for transgenic model: L1 et al., Cell Growth
and Differentiation 7, 3-11 (1996)."

BASE COUNT 217 a 145 c 174 g 166 t

ORIGIN

Query Match 32.3%; Score 557.8; DB 13; Length 702;
Best Local Similarity 89.4%; Pred. No. 1.1e-128;
Matches 623; Conservative 0; Mismatches 72; Indels 2; Gaps 2;

OY 293 AAAATATTAAGAAACAACTAAATAAAGTCCAGGAGAGATCATGAATCTAAGGAGA 352
|||||
DB 4 AAATTAAGAGAAACAACTAAGAAATGCTGGGAGAGATCATGAATCTTAAGAGGA 63
OY 353 GAATTAAGAAATGAGCTAAGAGCTCTGTTTCTTACAAAGTGAAGTGAAGAAATTA 412
|||||
DB 64 GATTTGAAGAAATGAGCTAAGAGCTCTGTTTCTTACAAAGTGAAGTGAAGAAATTA 123
OY 413 AGAATTTAGAGAAATGAGCTCCAAAGACATGACAGATGAATTTCTATCAGATTGGGAC 472
|||||
DB 124 AGCATTTAGAGAAATGAGCTCCAAAGACATGACAGATGAATTTCTTTGATTAGAGAC 183
OY 473 ATCATGAAGAGCTTAATAGAGGATCTTACTACTAGTCAACACAGATGGGCGAGGTG 532
|||||
DB 184 ACCATGAAGAGCTTAATAGAGGATCTTACTACTAGTCAACACAGATGGGCGAGGTG 243

QY 533 ATTGGCGTGAAGAGAGCCAAAGATCTGACAGAGCTGGTCCAGCGAGATTAATCATATC 592
|||||
DB 244 ATTGGCGTGAAGAGAGCCAAAGATCTGACAGAGCTGGTCCAGCGAGATTAATCATATC 303
593 TTCAGAAATCCAGAGAGCTGACAGAAAGCCAGAGAGCTAGTGTATATCAACAAAGGCT 652
|||||
DB 304 TCCGAATCTTAAAGAGCTGACAGAAAGCCAGAGAGCTGAGTGTATATCAATCAATAAAGGCT 363
QY 653 GTGGCTATGGCTGTACAGCTCCATCATGTAGTGTACTGCTTTATGATTCATATGGACCC 712
|||||
DB 364 GTGGCTATGGCTGTACAGCTCCATCATGTAGTGTACTGCTTTATGATTCATATGGACCC 423
QY 713 AACGAAGACTGCGCTGAGATCTCA-CAATGGAGCTGACCTAGTGGGAGGAGGAACCT 771
424 AACGAAGACTGCGCTGAGATCTCA-CAATGGAGCTGACCTAGTGGGAGGAGGAACCT 483
DB 772 GTGTTAGACCTGTAAAGAGAGCTGACAGAGAGATCTGGAGCTCCACTGGACATTTGG 831
484 GTGTTAGACCTGTAAAGAGAGCTGACAGAGAGATCTGGAGCTCCACTGGACATTTGG 543
QY 832 TCAGGTGAAGTAAAGCAAAATGTTCAAGTGTGAGTGTAGTGTAGTGTAGTGTAGTGT 891
|||||
DB 544 TCAGGTGAAGTAAAGCAAAATGTTCAAGTGTGAGTGTAGTGTAGTGTAGTGTAGTGT 603
QY 892 CATCTCGCTGCTGATATTTACCCCTGGCTGCTCCAGAGAGCTGACAGTGTAGTGTAGT 951
604 CATCTCG-6CTGCTTACTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662
QY 952 CGAGTCCATGATGATCTGACAGTGTGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 988
|||||
DB 663 AGAGTCCATGATGATCTGACAGTGTGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT 699

RESULT 11
LOCUS BG771085 715 bp mRNA linear EST 15-MAY-2001
DEFINITION 602719363p1 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:4839758 5',
mRNA sequence.
ACCESSION BG771085
VERSION BG771085.1 GI:14081738
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 715)
AUTHORS NIH-MGC http://mgi.nhl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgep@strausberg.nhl.gov
Tissue Procurement: DCTD/DBP
cDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L16M1669 row: n column: 15
High quality sequence stop: 665.
Location/Qualifiers
1, 715
FEATURES
Source
/organism="Homo sapiens"
/db_xref="Caxon:3606"
/clone="IMAGE:4839758"
/clone_lib="NIH_MGC_60"
/tissue_type="adipocarcinoma"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: prostate; Vector: pDMR-LIB (Clontech);
Site_1: SfiI (ggccgctcggcc); Site_2: SfiI (ggccatctggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'

adaptor sequence: 5'-CAGCGCATTTATGACC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGCGGAGATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH-MGC
Library."

BASE COUNT 215 a 160 c 151 g 189 t
ORIGIN
Query Match 32.2%; Score 557.2; DB 12; Length 715;
Best Local Similarity 89.8%; Pred. No. 1,5e+128;
Matches 643; Conservative 0; Mismatches 68; Indels 5; Gaps 4;
QY 975 GTGGTGGTATCCCACTTTGTCAAGTGTGATTTGCCCAACAACCTGGCTGGAAGA 1034
|||||
DB 1 GTGGTGGTGTCTCACTTTGTCAAAATGATTCCTCCCAACACCTTGGCTAGAAAAA 60
QY 1035 AATGAAAGAGCCCAAGAGCTAGGCTTCAAAATCCAGTATTTGAGTCCATTTAG 1094
|||||
DB 61 AATGAAAGAGCCCAAGAGCTTGGCTTCAAAATCCAGTATTTGAGTCCATTTAG 120
QY 1095 ACAGCAGACAAAGTGGAGAGGAGAGCTTCCATTCATTTGAGGAAATACAGGTGCA 1154
|||||
DB 121 ACAGCAGACAAAGTGGAGAGAGAGAGCTTCCATTCATTTGAGGAAATACAGGTGCA 180
QY 1155 CGTTGAAGAGACTTTCACCTTCTGCTGACAGATGCAAGTGTGATTAAGAGGTGTA 1214
|||||
DB 181 TGTGGAAGAACTTTTACCTTCTGCTGACAGATGCAAGTGTGATTAAGAGGTGTA 240
QY 1215 TTGGGCAAGATGAGCCCTGCTTTGTTAAAGAGGAAAGAAATACCCTGATTTA 1274
|||||
DB 241 TTGGGCAAGATGAGCCCTGCTTTGTTAAAGAGGAAAGAAATACCCTGATTTA 300
QY 1275 ATTATTAAGATGATGATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1334
|||||
DB 301 ATTATTAAGATGATGATCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
QY 1335 TTCACCTGGGGGTGATGATCTGATGATGATGATGATGATGATGATGATGATGATG 1394
|||||
DB 361 TTCACCTGGGGGTGATGATCTGATGATGATGATGATGATGATGATGATGATGATG 420
QY 1395 TACTTTTATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1454
|||||
DB 421 TACTTTTATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 1455 TGGCTCTGCAAGCTTCCGTTCTTTGATGATGATGATGATGATGATGATGATGATG 1514
|||||
DB 481 TGGCTCTGCAAGCTTCCGTTCTTTAGATGATGATGATGATGATGATGATGATGATG 540
QY 1515 CAACCAATTTGCAATTTATGCTGACCA-CTTCGAACTGAGAGAGAAATCCCATGAG 1573
|||||
DB 541 CAATCAATTTGCAATTTATGCTGACCA-CTTCGAACTGAGAGAGAAATCCCATGAG 600
QY 1574 CTGGAGATTA-TTATTTGCTGCTGAGAAATCACTGGAGTGGCTATTAAGGTGT -TA 1630
601 CTGGAGATTA-TTATTTGCTGCTGAGAAATCACTGGAGTGGCTATTAAGGTGTCCA 660
QY 1631 ACAGAAAGCTGAGAGAGGAGGCTTATTCCTCTACAAAGTTCAGAGAGAGATA 1686
|||||
DB 661 CCGGACAAATGGAGAGGAGGAGGCTTATTCCT-CTACAAAGTTCAGAGAGAGATA 715

RESULT 12
LOCUS B0552954 656 bp mRNA linear EST 20-JUN-2002
DEFINITION H4019B10-5 N1A Mouse 7.4K cDNA clone set Mus musculus cDNA clone
H4019B10 5', mRNA sequence.
ACCESSION B0552954
VERSION B0552954.1 GI:21453842
KEYWORDS EST.
SOURCE house mouse.


```

ORGANISM      Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
                1 (bases 1 to 656)
                Vanburen, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin
                , P.R., Stagg, C.A., Bassey, U., Aliba, K., Hamatani, F., Kargul, G.J.,
                Luo, A.G. and Ko, M.S.H.
TITLE          Assembly, verification, and initial annotation of NIA 7.4K mouse
                cDNA clone set
JOURNAL        Unpublished (2002)
COMMENT        Other ESTs: H4019B10-3
                Contact: Yong Qian
                Laboratory of Genetics
                National Institute on Aging/National Institutes of Health
                333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
                Email: cdna@igsun.grc.nia.nih.gov
                This clone set has been freely distributed to the community. Please
                visit http://igsun.grc.nia.nih.gov/cDNA/NIA_7_4K.html for details.
                Plate: H4019 row: B column: 10
                Seq primer: -21M13 Reverse
                High quality sequence stop: 656
                POLYA-No.

FEATURES
source        Location/Qualifiers
                1..656
                /organism="Mus musculus"
                /strain="C57Bl/6"
                /db_xref="niaEST:H4019B10-5"
                /db_xref="taxon:10090"
                /clone="H4019B10"
                /clone_1lib="NIA Mouse 7.4K cDNA Clone Set"
                /sex="mixed"
                /dev_stage="mixed"
                /lab_host="DH10B"
                /note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This
                clone is among a rearranged set of 7,407 clones from more
                than 20 cDNA libraries."

BASE COUNT    190 a 153 c 145 g 168 t

Query Match   32.2%; Score 555.8; DB 14; Length 656;
Best Local Similarity 90.5%; Pred. No. 3.3e-128;
Matches 593; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Oy 786 AAGTGAGACGTGCAGACAGATCTGGACATCCATCGATGGATGGTCAAGTAA 845
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 1 AAGTGAGACATGTACAGACAGATCTGGCCCTCTCAGCTGGACAGCTGGTCAAGTAA 60

Oy 846 GGCACAAATGTTCAAGTGGTGAAGTCCCATTTGACAGAGTGTTCCTGCTCC 905
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 61 TGCACAAACATTCAGATGCTGAGCTCCCATTTGACAGAGCTCCATCTCGGCTCC 120

Oy 906 ATATTTACCCGTGGTGTCCAGACAGACCTTTCAGATTCGATTCAGATGGATG 965
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 121 TTTACTTACCACTGGCTGTTTCCAGAAAGACCTTGCAGACCGACTCTTAAGATCCATG 180

Oy 966 TCCGTGAGTGTGGTGGTATCCCATTTGTCAAGTACTTGATTCGCCCAACCCCTGGCT 1025
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 181 CCTGCAAGTGTGGTGGTGTCCCAATTTGTCAATTAATGATTCGTCACAAACCTTGGCT 240

Oy 1026 GGAAGAAAGAAATGAAAGAGGCCCAAGAAAGCTAGGCTTCAAAATCCAGTTATTTGAG 1085
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 241 GGAAGAAAGAAATGAAAGAGGCCCAAGAAAGCTTGGCTTCAAAATCCAGTTATTTGAG 300

Oy 1086 CCATGTTAGACGACAGACAAAGTGGAGGAGGAGAGGAGGAGGAGGAGGAGGAGGAG 1145
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 301 CCATGTTAGACGACAGACAAAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360

Oy 1146 CAGCGTGCAGCTTGAAGAAGACTTTCAGCTTCTTGGCTCGAGAGATCAAGTGGATTA 1205
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 361 CAGGTGACAGCTTGAAGAAGACTTTCAGCTTCTTGGCTCGAGAGAGAGATCAAGTGGATTA 420

Oy 1206 AAGGGTGTATTTGGCCACAGATGACCTGCTTTGTTAAAGAGGCAAAACAAAGTACCC 1265
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||

```

```

Db 421 AAGATATATCTGGCTAGCTGATGATCCTACTTGTAAAGAGCAAGAAAGTACTC 480
Oy 1266 CAGTTATGAATTTATAGTATACCTATCTCTGTGAGTGGAGTACATATGATA 1325
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 481 CAATTATGAATTTATAGTATACCTATCTCTGTGAGTGGAGTACATATGATA 540

Oy 1326 TACAGAAATTCACCTCGGGGTGTATCCGTGATATACACTTTCTGCCAGCAGACT 1385
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 541 CACAGAAATTCACCTCGGGGTGTATCCGTGATATACACTTTCTGCCAGCAGACT 600

Oy 1386 CCTAGTGTACTTTTTCATCGCAGGCTGTATACAGTTGCTTATGAATATGCA 1440
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 601 TCTAGTGTACTTTTTCATCGCAGGCTGTATACAGTTGCTTATGAATATGCA 655

RESULT 13
LOCUS      B0731540
DEFINITION B0731540 900 bp mRNA linear EST 16-JUL-2002
AGENCOURT.8214839 NICHD XGC Embd4 Xenopus laevis cDNA clone
IMAGE:4724851 5', mRNA sequence.

ACCESSION  B0731540
VERSION    B0731540.1 GI:21870437
KEYWORDS   EST.
SOURCE     African clawed frog.
ORGANISM   Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Piploidea; Pipidae;
            Xenopodinae; Xenopus.
            1 (bases 1 to 900)
REFERENCE  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE     Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dr. Igor David
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/dbfp/image/image.html
            Plate: LLAM10581 row: 3 column: 20
            High quality sequence stop: 607.

FEATURES
source        Location/Qualifiers
                1..900
                /organism="Xenopus laevis"
                /db_xref="taxon:8355"
                /clone="IMAGE:4724851"
                /clone_1lib="NICHD XGC Embd4"
                /dev_stage="embryo, stage 31-32"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: whole embryo; Vector: pCMV-Sport6; Site 1:
                NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
                dt. Average insert size 2.1 kb. Constructed by Life
                Technologies. Note: This is a Xenopus Gene Collection (XGC
                ) library."

BASE COUNT    242 a 223 c 219 g 216 t

Query Match   32.0%; Score 552.4; DB 14; Length 900;
Best Local Similarity 77.6%; Pred. No. 2.5e-127;
Matches 681; Conservative 0; Mismatches 196; Indels 1; Gaps 1;

Oy 492 GACGATCTATCTACTCTACGTCACATGAGGAGGAGGAGGAGGAGGAGGAGGAG 551
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 1 GACGATCTCTACTCTACGTCACATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 60

Oy 552 CAAAGATTCAGACAGAGCTGTCCAGGAGAAATACATATCTTCACAAATCCCCAAGAGCT 611
      |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||  |||||||
Db 61 CAAAGATTCAGACAGATCTGTACAAAGAAATACATATCTTCACAAACCCCAAGAGATT 120

Oy 612 CAGCAAGGCCAGAGACTAGTGTATATCAACAAAGGCTGTGGCTATGGCTGACCT 671

```


